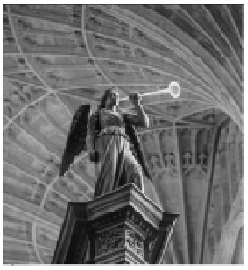


Steganography

Information Hiding

Today..

- What is steganography?
- Traces in history
- Modern day applications
- Steganographic model
- Steganography in MP3
- Steganography in digital images
- Steganography in Text

[illegible]

STEGANOGRAPHY

STEGANOGRAPHY

[illegible]

STEGANOGRAPHY

cover * message * stego

STEGANOGRAPHY

What is Steganography?

- Embedding information in given media without making any visible changes to it.
- It replaces unneeded/redundant bits in image, sound, and text files with secret data.
- Instead of protecting data the way encryption does, steganography hides the very existence of the data.

Steganography and Steganalysis

- Steganography
  - Goal – hide an embedded file within a cover file such that embedded file’s existence is concealed
  - Result is called stego file
  - Substitution (least significant bit), transform, spread spectrum, cover generation, etc
- Steganalysis
  - Goals – detection, disabling, extraction, confusion of steganography
  - Visible detection, filtering, statistics, etc

Ref: Katz, West, John, Frid, Fari

Problem formulation

- Problem:
  - Alice and Bob are in male/female prisons and want to communicate to make an escape plan. Willie warden would let them communicate but would monitor the communication.
  - A solution needs to be found out such that the communication would seem to be innocent to person who is not aware that “something lies beneath it”.

Traces in history

- Existed in different forms and media
- Tatoos
- Dots on top of *‘i’* and *‘j’*
- Deliberate misspellings or Error

Modern day applications

- Avoid third party snooping
- Security reinforcement layer to cryptography
- Hiding copyright info: digital watermarks and fingerprinting (growing due to web piracy)
- Data encapsulation : data and still images

Modern day applications

- medical doctors combine explanatory or serious information within X-ray images
- communications for codes self-error correcting; corrective audio or image data in case corruption
- copyright protection is to protect the cover medium from claiming its credit be others

Steganography & Cryptography

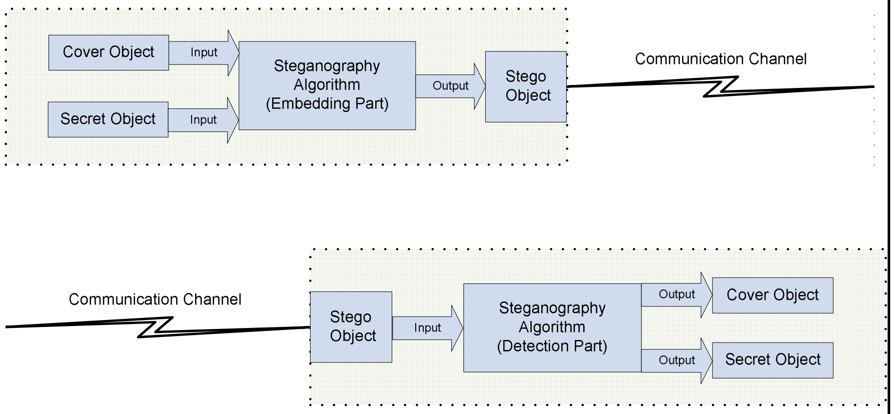
- have overlapping usages in the information hiding.
- Steganography security hides the knowledge that there is information in the cover medium.
- cryptography revels this knowledge but encodes the data as cipher-text and disputes decoding it without permission.
- cryptography concentrate the challenge on the decoding process while steganography adds the search of detecting if there is hidden information or not.

Steganography & Watermarking

- Have overlapping usages in the info hiding & intellectual rights issues
- Watermarking is different from steganography in its main goal.
- Watermarking aim is to protect the cover medium from any modification with no real emphasis on secrecy.
- Watermarking can be observed as steganography that is concentrating on high robustness and very low or almost no security.

Steganographic model

Basic Steganography System



Cover Object

Input

Secret Object

Input

Steganography Algorithm (Embedding Part)

Output

Stego Object

Communication Channel

Communication Channel

Stego Object

Input

Steganography Algorithm (Detection Part)

Output

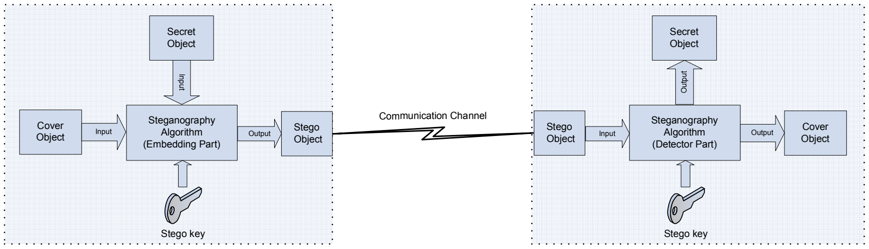
Cover Object

Output

Secret Object

Steganographic model

PRNG-Key Steganography System



Secret Object

Input

Cover Object

Input

Steganography Algorithm (Embedding Part)

Output

Stego Object

Stego key

Communication Channel

Secret Object

Output

Stego Object

Input

Steganography Algorithm (Detector Part)

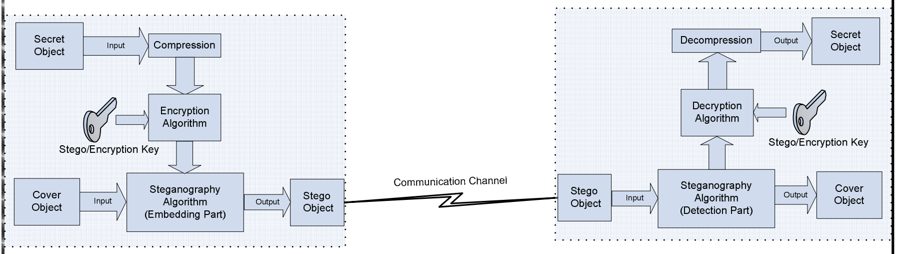
Output

Cover Object

Stego key

Steganographic model

Advanced Steganography System


Secret Object
Input
Compression
Encryption Algorithm
Stego/Encryption Key
Cover Object
Input
Steganography Algorithm (Embedding Part)
Output
Stego Object
Communication Channel
Stego Object
Input
Steganography Algorithm (Detection Part)
Output
Cover Object
Decryption Algorithm
Stego/Encryption Key
Decompression
Output
Secret Object

Stego-system criteria

- Cover data should not be significantly modified i.e. perceptible to human perception system
- The embedded data should be directly encoded in the cover & not in wrapper or header
- Embedded data should be immune to modifications to cover
- Distortion cannot be eliminated so error-correcting codes need to be included whenever required

Steganography main aspects and usefulness

- Security
  - ability of an eavesdropper to figure the hidden information easily
- Capacity
  - amount of data bits that can be hidden in the cover medium
- Robustness
  - resist possibility of modifying or destroying the unseen data

Stego cover media

- Any electronic file type:
  - Audio
  - Image
  - Text
  - Video

Steganography in Audio

- Audio company publishes Audio products in mp3 and publishes over internet.
- Some people take these mp3 files and publish under their own name.
- Case goes to court.
- The Audio company needs to prove that the material which is exhibit is indeed the one they published.
- They need a hidden copyright.

A Picture is Worth 1,024 Words

- Images have a lot of data:
  - Spam Mail: 1,108 Bytes
  - Picture of Dave: 152 KB
- Ways to hide information
  - Least significant bit (BMP)
  - Palette Shifts (GIF)
  - Discrete Cosine Transforms (JPG)

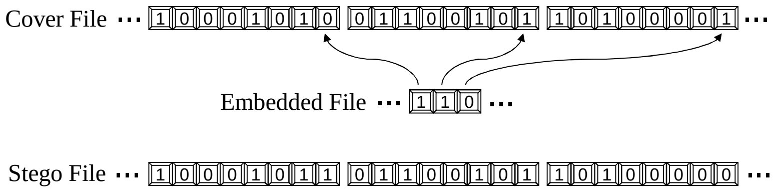
Steganography in images

Way images are stored:

- Array of numbers representing RGB values for each pixel
- Common images are in 8-bit/pixel and 24-bit/pixel format.
- 24-bit images have lot of space for storage but are huge and invite compression
- 8-bits are good options.
- Proper selection of cover image is important.
- Best candidates: gray scale images ………RGB images

Least significant bit (LSB) substitution Steganography

- Easy to understand and implement
- Used in many available stego tools



Cover File ··· 1 0 0 0 1 0 1 0 0 1 1 0 0 1 0 1 1 0 1 0 0 0 0 1 ···

Embedded File ··· 1 1 0 ···

Stego File ··· 1 0 0 0 1 0 1 1 0 1 1 0 0 1 0 1 1 0 1 0 0 0 0 0 ···

Least Significant Bit method

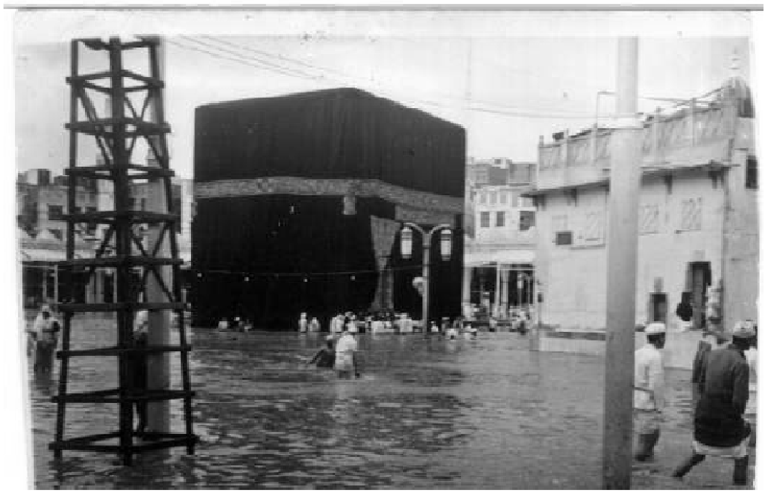
- Consider a 24 bit picture
- Data to be inserted: character ‘*A*’: (10000011)
- Host pixels: 3 pixel will be used to store one character of 8-bits
- The pixels which would be selected for holding the data are chosen on the basis of the key which can be a random number.
- Ex: 00100111 11101001 11001000
  00100111 11001000 11101001
  11001000 00100111 11101001

Embedding ‘*A*’

0010011**1** 1110100***0*** 1100100**0**
0010011***0*** 1100100**0** 1110100***0***
1100100**1** 0010011**1** 11101001

- According to researchers on an average only 50% of the pixels actually change from 0-1 or 1-0.

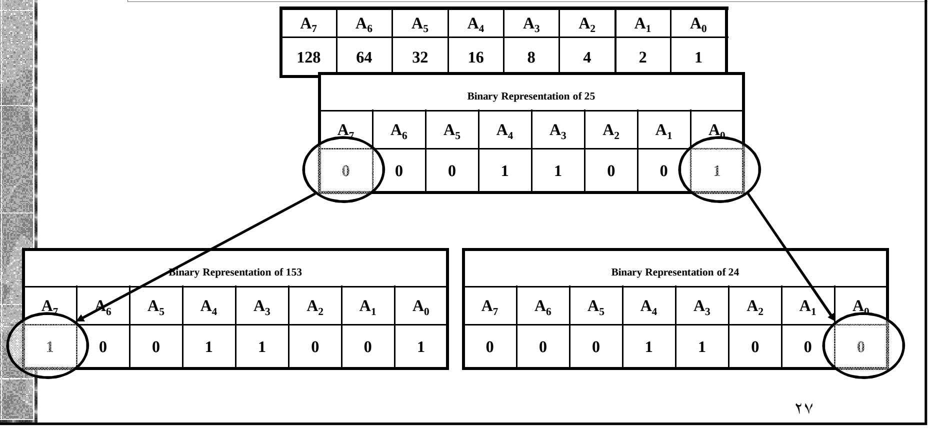
What’s behind this pic?



gray-scale

(Bit Planes)

- An Image is a 2D array of pixels
- Each pixel (in gray-scale) is 8 bits
- Each bit carries information
  - Not all bits carry the same ***amount***



A_7	A_6	A_5	A_4	A_3	A_2	A_1	A_0
128	64	32	16	8	4	2	1

Binary Representation of 25

A_7	A_6	A_5	A_4	A_3	A_2	A_1	A_0
0	0	0	1	1	0	0	1

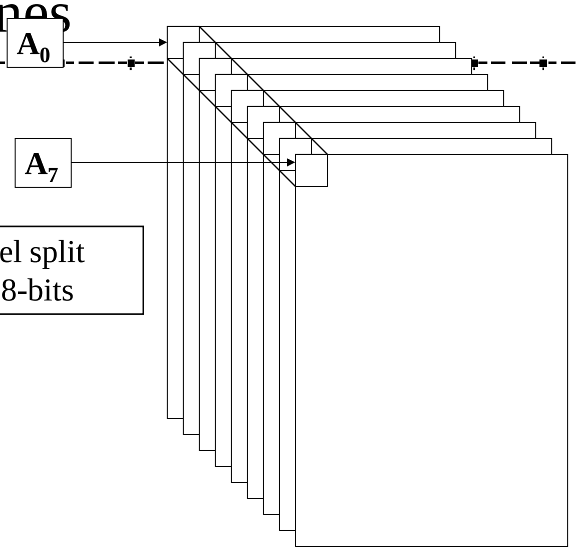
Binary Representation of 153

A_7	A_6	A_5	A_4	A_3	A_2	A_1	A_0
1	0	0	1	1	0	0	1

Binary Representation of 24

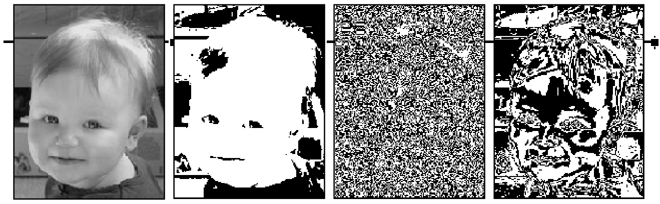
A_7	A_6	A_5	A_4	A_3	A_2	A_1	A_0
0	0	0	1	1	0	0	0

Bit Planes



One pixel split
into it’s 8-bits

Bit Planes



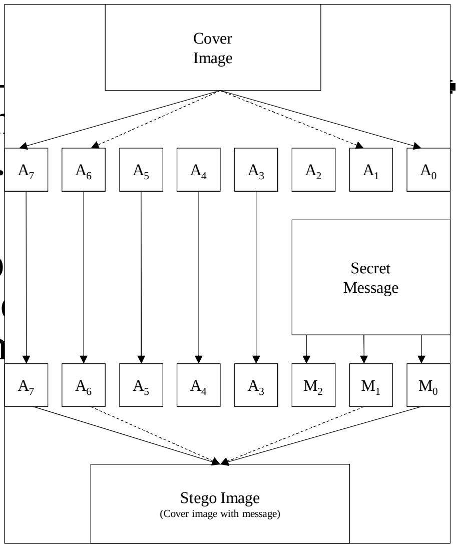
(a) (b) (c) (d)

White indicates an ON bit and black an OFF bit

(a) 8-bit grayscale source image
(b) Most significant bit plane (*a7)* of the source
(c) Least significant bit plane (*a0)* of the source
(d) Bit plane *a4* of the source image

The Big Idea

(LSB embedding)

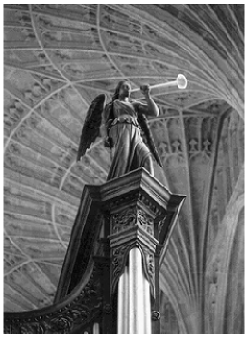

Cover Image
A_7
A_6
A_5
A_4
A_3
A_2
A_1
A_0
Secret Message
A_7
A_6
A_5
A_4
A_3
M_2
M_1
M_0
Stego Image
(Cover image with message)

- An image contain can be perceived.
- Replace the “imp noticeable) bits” (bits of a “secret n

+

TOP SECRET

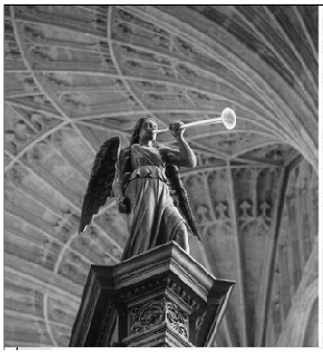
=



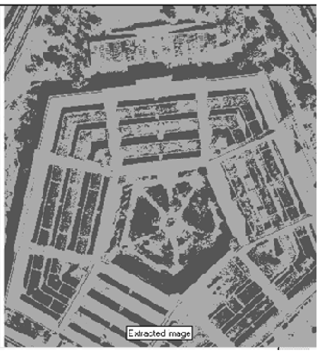
8-bit (256 grayscale)
images.

Example: Copyright Fabian A.P. Petitcolas,
Computer Laboratory, University of Cambridge
http://www.cl.cam.ac.uk/~fapp2/steganography/image_downgrading/

Sacrificing 2 bits of cover to carry 2 bits of secret image

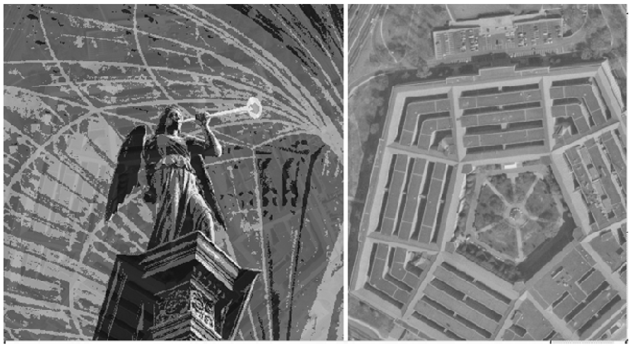


Original Image



Extracted Image

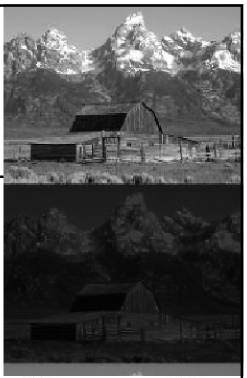
Sacrificing 5 bits of cover to carry 5 bits of secret image



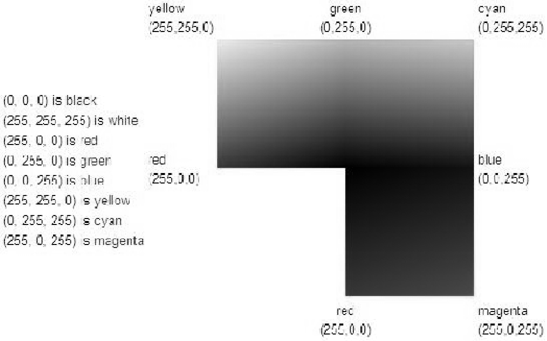
Original Image Extracted Image

RGB Image Steganography

pictures from wikipidea (http://en.wikipedia.org/wiki/RGB_color_model)



- RGB Pixel = 3 channels: Red, Green, Blue
- **RGBA** = Red, Green, Blue, Alpha. RGB with Alpha = transparency.



(0, 0, 0) is black
(255, 255, 255) is white
(255, 0, 0) is red
(0, 255, 0) is green
(0, 0, 255) is blue
(255, 255, 0) is yellow
(0, 255, 255) is cyan
(255, 0, 255) is magenta

yellow
(255,255,0)
green
(0,255,0)
cyan
(0,255,255)
red
(255,0,0)
blue
(0,0,255)
red
(255,0,0)
magenta
(255,0,255)

RGB Image Steganography

- Single channel hiding
  - Hide & Seek (Provos & Honeyman 2003)
  - Stego-1bit, Stego-2bits, Stego-3bits & Stego-4bits
- Multi channel hiding
  - Stego Color Cycle (SCC)
  - Pseudorandom Number Generator (PRNG)
  - S-Tools
- (Curran & Bailey 2006)
- Pixel indicator technique

Stego Color Cycle (SCC)

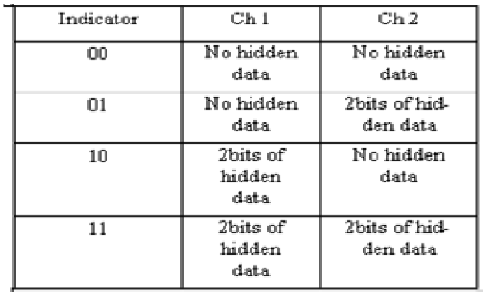
- RGB images
- Hide a bit/pixel
- One channel is utilized
- Hiding channels cycles in a sequence
- Improvement = hide more than a bit

Steganography

Using Pixel Indicator technique for better Steganography

Pixel Indicator Technique

- RGB images are of 24-bits per pixels.
- Use of LSB bits of one of the channels as indicator for data existence in the other two channels.
- The 2 LSB of indicator are based on Image nature.
- First indicator is chosen based on image length value property



Indicator	Ch 1	Ch 2
00	No hidden data	No hidden data
01	No hidden data	2bits of hidden data
10	2bits of hidden data	No hidden data
11	2bits of hidden data	2bits of hidden data

Indicator sequence

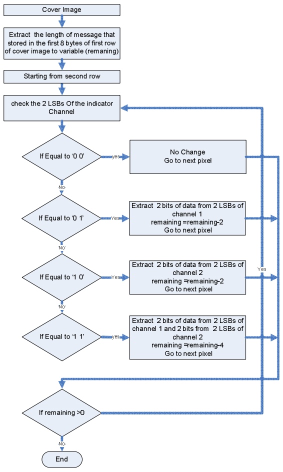
- First indicator channel

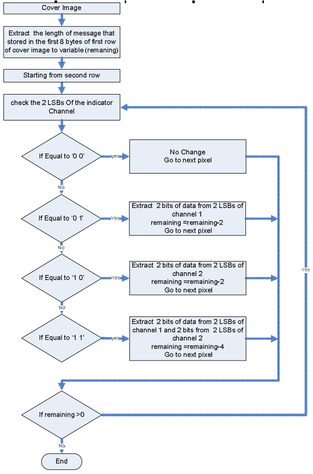
Type of length (N) of secret message	I Level Selection Select indicator channel, first element of sequence	II Level Selection Binary N parity-bit	
		Odd Parity	**Even Parity**
Even	R	GB	BG
Prime	B	RG	GR
Else	G	RB	BR

- Steganography sequence:

  -RGB-RBG-GBR-GRB-BRG-BGR-

Pixel Indicator Technique Process


Cover Image
Extract the length of message that stored in the first 8 bytes of first row of cover image to variable (remaning)
Starting from second row
check the 2 LSBs Of the indicator Channel
If Equal to '0 0'
yes
No Change
Go to next pixel
No
If Equal to '0 1'
Yes
Extract 2 bits of data from 2 LSBs of channel 1
remaining =remaining-2
Go to next pixel
No
If Equal to '1 0'
Yes
Extract 2 bits of data from 2 LSBs of channel 2
remaining =remaining-2
Go to next pixel
No
If Equal to '1 1'
yes
Extract 2 bits of data from 2 LSBs of channel 1 and 2 bits from 2 LSBs of channel 2
remaining =remaining-4
Go to next pixel
Yes
If remaining >0
No
End


Cover Image
Extract the length of message that stored in the first 8 bytes of first row of cover image to variable (remaning)
Starting from second row
check the 2 LSBs Of the indicator Channel
If Equal to '0 0'
yes
No Change
Go to next pixel
No
If Equal to '0 1'
Yes
Extract 2 bits of data from 2 LSBs of channel 1
remaining =remaining-2
Go to next pixel
No
If Equal to '1 0'
Yes
Extract 2 bits of data from 2 LSBs of channel 2
remaining =remaining-2
Go to next pixel
No
If Equal to '1 1'
yes
Extract 2 bits of data from 2 LSBs of channel 1 and 2 bits from 2 LSBs of channel 2
remaining =remaining-4
Go to next pixel
Yes
If remaining >0
No
End

Testing And Results

- Image size = 512 X 384 = 196608 pixels
- Secret text =11,733 characters length = 93,864 bits
- The test performed hiding data using

  1 bit, 2 bits, 3 bits, 4 bits, 5bits.
- Histogram for each channel in each run was drawn.
- The number of pixels required each time was recorded.

Testing And Results (2bits)

- Using 2 bits as hidden data:

Original (Cover)+ secret = Stega.obj.

Testing And Results (3bits)

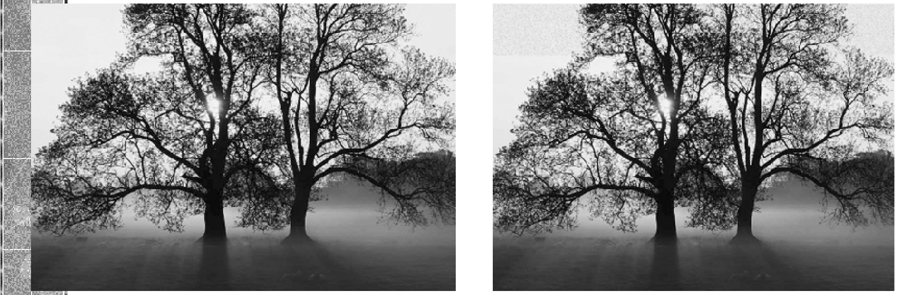
- Using 3 bits as hidden data:

Original (Cover)+ secret = Stega.obj.

Testing And Results (4bits)

- Using 4 bits as hidden data:

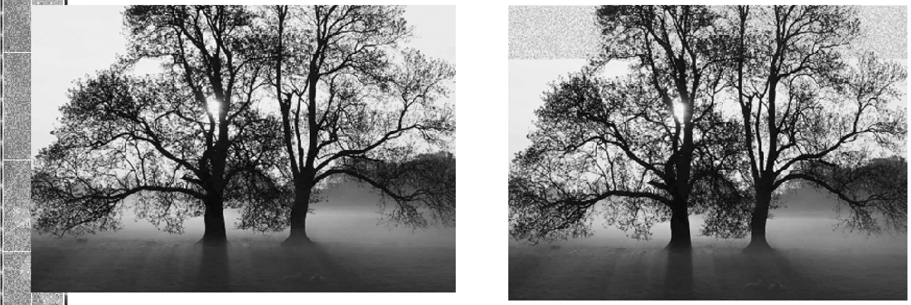
Original (Cover)+ secret = Stega.obj.



Testing And Results (5bits)

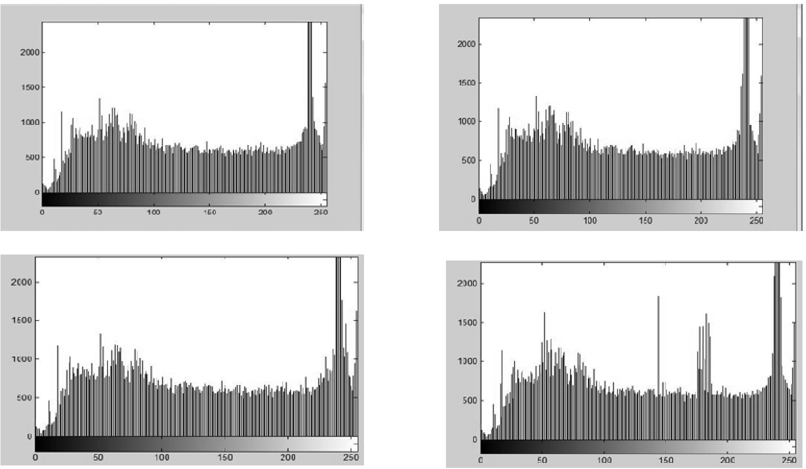
- Using 5 bits as hidden data:

Original (Cover)+ secret = Stega.obj.



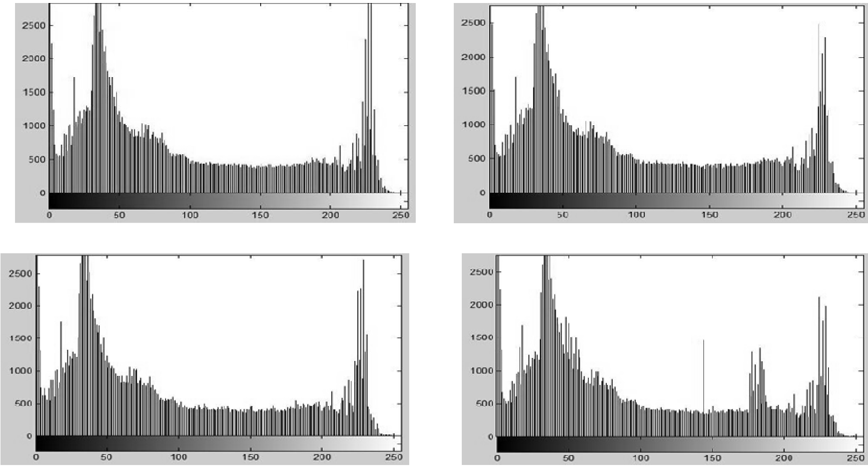
Testing And Results (Red Channel)

Original image, 2-bits, 4-bits and 5-bits:



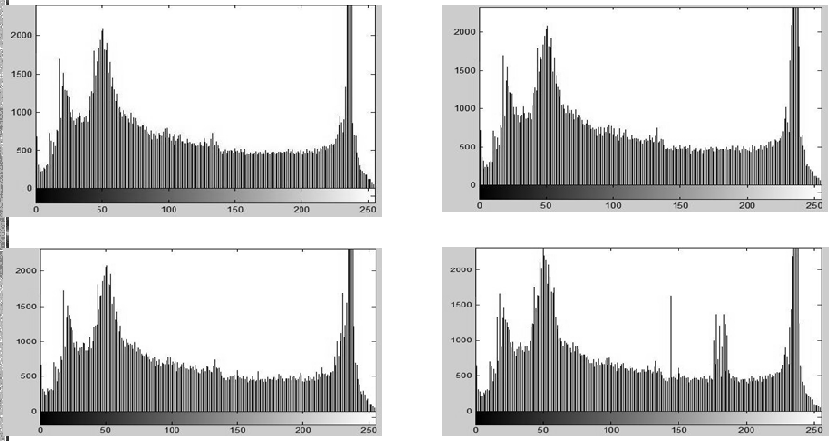
Testing And Results (Blue Channel)

- Original image, 2-bits, 4-bits and 5-bits):

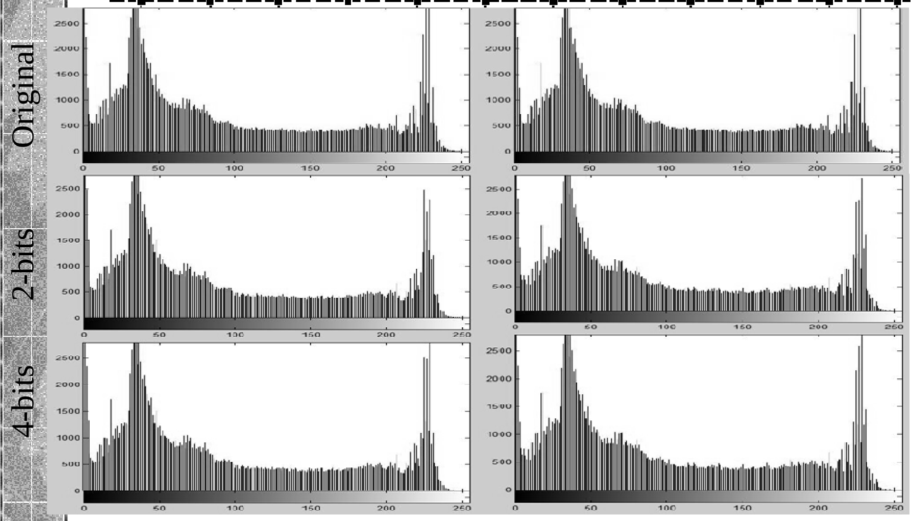


Testing And Results (Green Channel)

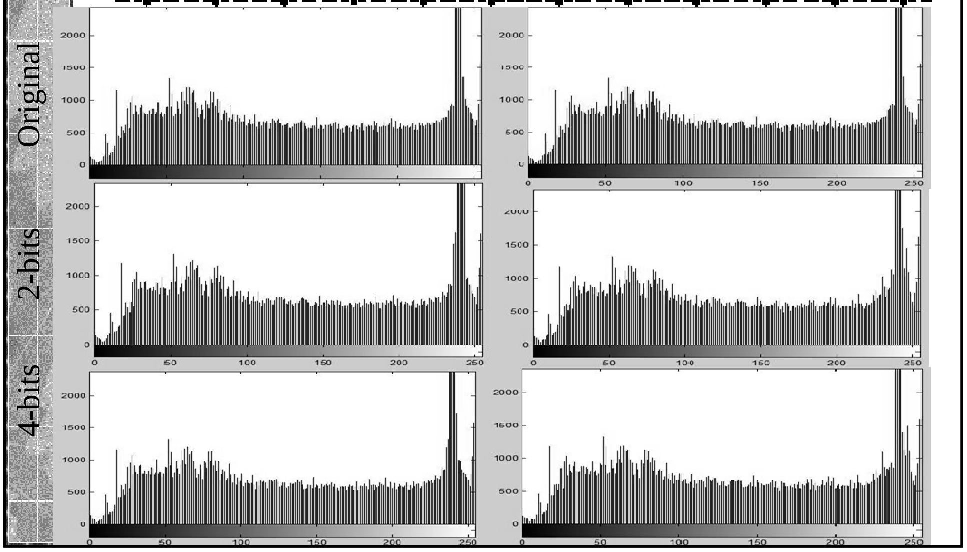
- Original image, 2-bits, 4-bits and 5-bits:



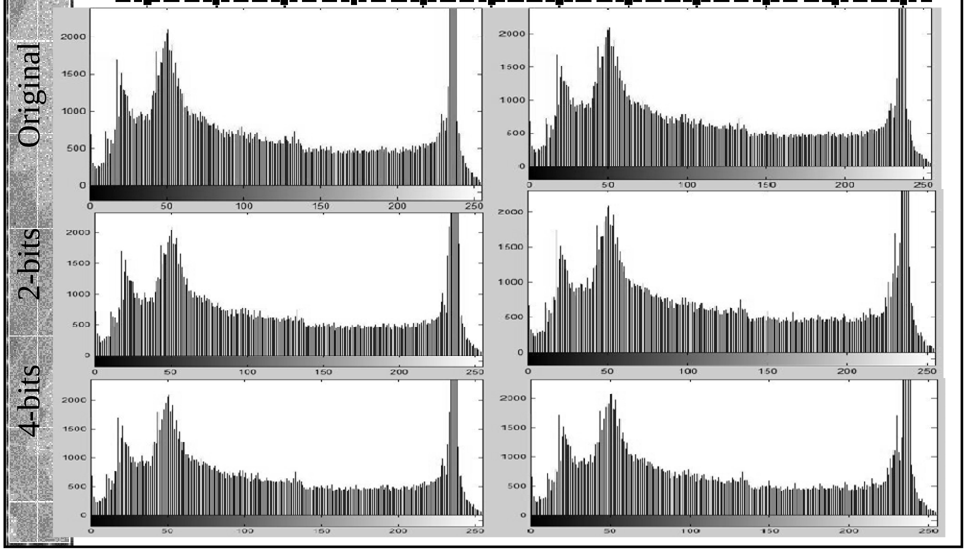
Pixel Indicator Technique Vs Stego Color Cycle (Blue Channel)



Pixel Indicator Technique Vs Stego Color Cycle (RED Channel)



Pixel Indicator Technique Vs Stego Color Cycle (Green Channel)



Spaces

- By adding extra white-spaces between words, or at end of lines or paragraph of the text
- What are the strengths & weaknesses?
- Does not reveal secrecy to the normal reader ➔ high security
- Cannot hide too much information ➔ low capacity
- Electronic text editors automatically remove extra white-spaces ➔ low robustness

---

Dear Friend , Your email address has been submitted to us indicating your interest in our newsletter . If you no longer wish to receive our publications simply reply with a Subject: of "REMOVE" and you will immediately be removed from our club . This mail is being sent in compliance with Senate bill 2116 , Title 7 , Section 301 ! Do NOT confuse us with Internet scam artists . Why work for somebody else when you can become rich within 13 DAYS ! Have you ever noticed the baby boomers are more demanding than their parents and nobody is getting any younger ! Well, now is your chance to capitalize on this . WE will help YOU process your orders within seconds plus deliver goods right to the customer's doorstep . The best thing about our system is that it is absolutely risk free for you ! But don't believe us ! Mr Ames who resides in Delaware tried us and says "My only problem now is where to park all my cars" ! This offer is 100% legal ! If not for you then for your **loved ones** – act now ! Sign up a friend and you'll get a discount of 50% . Thank-you for your serious consideration of our offer !

Give us an A-

Dear Friend , Your email address has been submitted to us indicating your interest in our newsletter . If you no longer wish to receive our publications simply reply with a Subject: of "REMOVE" and you will immediately be removed from our club . This mail is being sent in compliance with Senate bill 2116 , Title 7 , Section 301 ! Do NOT confuse us with Internet scam artists . Why work for somebody else when you can become rich within 13 DAYS ! Have you ever noticed the baby boomers are more demanding than their parents and nobody is getting any younger ! Well, now is your chance to capitalize on this . WE will help YOU process your orders within seconds plus deliver goods right to the customer's doorstep . The best thing about our system is that it is absolutely risk free for you ! But don't believe us ! Mr Ames who resides in Delaware tried us and says "My only problem now is where to park all my cars" ! This offer is 100% legal ! If not for you then for your **LOVED ONES** – act now ! Sign up a friend and you'll get a discount of 50% . Thank-you for your serious consideration of our offer !

Give us an A+

Abbreviation Steganography using letter cases

Semantic Method

- Semantic method proposes using synonyms of words for certain words as for hiding information in the text.
- However, this method may alter the meaning of the text which will change the intended hidden information.

Character Feature Steganography

- Changes some of the features of the text characters.
- Example, the most significant bits of some characters are extended to hold bits of the hidden information.
- Character steganography can hold a large quantity of secret information without making normal readers aware of the existence of such information in the text.

Arabic Text Steganography

Dotted letters (pointed letters)

Extensions (Kashida) & dotted letters

Diacritics

Arabic Based Steganography

Arabic language is the largest living member of the Semitic language family in terms of speakers. (270 million speakers).

It contains 28 alphabet characters; 15 of which have points.

اللُّغَةُ

العَرَبِيْةُ

Characters with no points	Characters with one point	Characters with two points	Characters with three points
أ ح د ر س ص ط ع ك ل م هـ و	ب ج خ ذ ز ض ظ غ ف ن	ت ق ي	ث ش

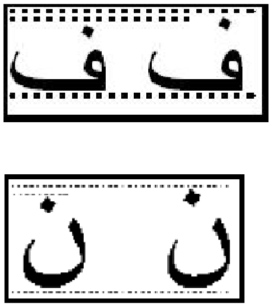
Fig. 2: Arabic Alphabet

Characteristics of the Arabic Script

- The Arabic alphabet has Semitic origins derived from the Aramaic writing system .
- Little has been proposed for Arabic script steganography.
- Two inherent properties of Arabic writing: dots & connectability.
- Arabic basic alphabet of 28 letters, 15 have from one to three points, four letters can have a Hamzah, and one, ALEF, can be adorned by the elongation stroke, the Maddah

Pointed Letters

- Shirali-Shahreza, 2006

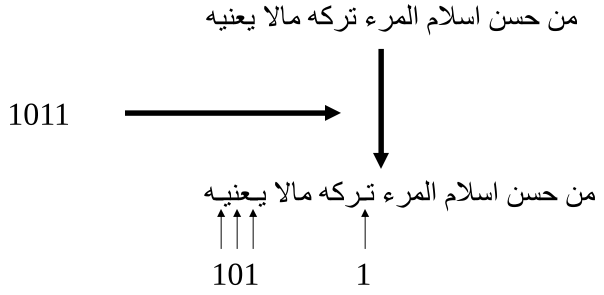


un-pointed letters = 13	pointed letters = 15
ا ح د ر س ص ط ع ك ل م هـ و	ب ت ث ج خ ذ ز ش ض ظ غ ف ق ن ي

- Drawback: robustness

Pointed Letters & Extensions (Kashida)

- Steganography example adding extensions after pointed letters.


من حسن اسلام المرء تركه مالا يعنيه

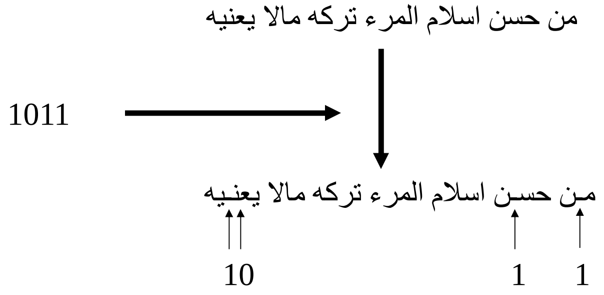
1011

من حسن اسلام المرء تـركه مالا يـعنيـه

101 1

Pointed Letters & Extensions (Kashida)

- Steganography example adding extensions before pointed letters.


من حسن اسلام المرء تركه مالا يعنيه

1011

مـن حسـن اسلام المرء تركه مالا يعـنيه

10 1 1

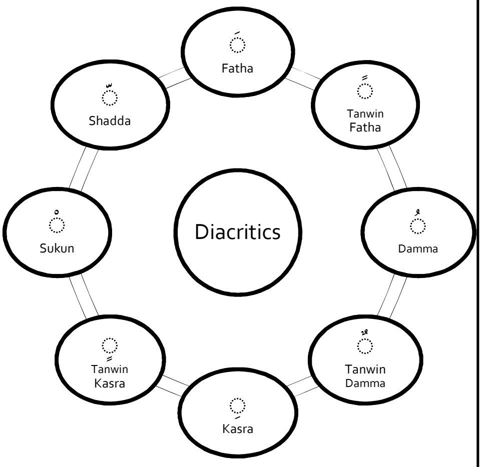
ARABIC DIACRITICS حركات BASED STEGANOGRAPHY

Diacritics (Harakat – حركات)

Arabic language uses eight symbols as diacritical marks.

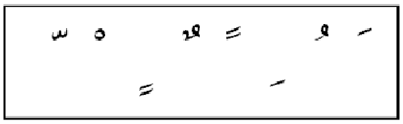
It is used to alter the pronunciation of a phoneme or to distinguish between words of similar spelling.

The use of diacritics in the text is optional in written Standard Arabic.


Diacritics
◌َ Fatha
◌ً Tanwin Fatha
◌ُ Damma
◌ٌ Tanwin Damma
◌ِ Kasra
◌ٍ Tanwin Kasra
◌ْ Sukun
◌ّ Shadda

Arabic Diacritics

Background on Arabic Diacritic Marks



- **Arabic diacritics decorate letters in Arabic script and modify their pronunciation**
- **Vowels occur pretty frequently in languages**
- **Read the English sentence by deducing vowel diacritics and feel the difference**

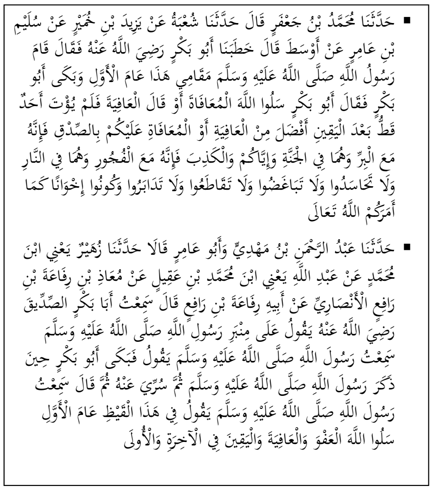
Just to feel the task, read the following English sentence:
"jst t fl th tsk, rd th fllwng nglsh sntnc"

Statistics for Diacritics

First we needed to find the average occurrences of diacritics in a fully diacritized Arabic document.

Then we needed to compare these occurrences to find the best embedding technique available.

Both ambiguity and capacity are important factors to consider.



- حَدَّثَنَا مُحَمَّدُ بْنُ جَعْفَرٍ قَالَ حَدَّثَنَا شُعْبَةُ عَنْ يَزِيدَ بْنِ خُمَيْرٍ عَنْ سُلَيْمِ بْنِ عَامِرٍ عَنْ أَوْسَطَ قَالَ خَطَبَنَا أَبُو بَكْرٍ رَضِيَ اللَّهُ عَنْهُ فَقَالَ قَامَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ مَقَامِي هَذَا عَامَ الْأَوَّلِ وَبَكَى أَبُو بَكْرٍ فَقَالَ أَبُو بَكْرٍ سَلُوا اللَّهَ الْمُعَافَاةَ أَوْ قَالَ الْعَافِيَةَ فَلَمْ يُؤْتَ أَحَدٌ قَطُّ بَعْدَ الْيَقِينِ أَفْضَلَ مِنْ الْعَافِيَةِ أَوْ الْمُعَافَاةِ عَلَيْكُمْ بِالصِّدْقِ فَإِنَّهُ مَعَ الْبِرِّ وَهُمَا فِي الْجَنَّةِ وَإِيَّاكُمْ وَالْكَذِبَ فَإِنَّهُ مَعَ الْفُجُورِ وَهُمَا فِي النَّارِ وَلَا تَحَاسَدُوا وَلَا تَبَاغَضُوا وَلَا تَقَاطَعُوا وَلَا تَدَابَرُوا وَكُونُوا إِخْوَانًا كَمَا أَمَرَكُمْ اللَّهُ تَعَالَى
- حَدَّثَنَا عَبْدُ الرَّحْمَنِ بْنُ مَهْدِيٍّ وَأَبُو عَامِرٍ قَالَا حَدَّثَنَا زُهَيْرٌ يَعْنِي ابْنَ مُحَمَّدٍ عَنْ عَبْدِ اللَّهِ يَعْنِي ابْنَ مُحَمَّدِ بْنِ عَقِيلٍ عَنْ مُعَاذِ بْنِ رِفَاعَةَ بْنِ رَافِعٍ الْأَنْصَارِيِّ عَنْ أَبِيهِ رِفَاعَةَ بْنِ رَافِعٍ قَالَ سَمِعْتُ أَبَا بَكْرٍ الصِّدِّيقَ رَضِيَ اللَّهُ عَنْهُ يَقُولُ عَلَى مِنْبَرِ رَسُولِ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ سَمِعْتُ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ يَقُولُ فَبَكَى أَبُو بَكْرٍ حِينَ ذَكَرَ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ ثُمَّ سُرِّيَ عَنْهُ ثُمَّ قَالَ سَمِعْتُ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ يَقُولُ فِي هَذَا الْقَيْظِ عَامَ الْأَوَّلِ سَلُوا اللَّهَ الْعَفْوَ وَالْعَافِيَةَ وَالْيَقِينَ فِي الْآخِرَةِ وَالْأُولَى

Fig. 6: Sample for diactrized Arabic text

Statistics for sample file

Characters ١٠٤١٤٤٣

Tanwin Fatha ٢٥١٨

Tanwin Damma 2768

Tanwin Kasra 6317

Sukun 59033

Shadda 37282

Kasra 59982

Damma 48065

Fatha ٢٠٨٦٥٤

Rest of ch. ٦١٧٤٠٤

Total Harakat ٤٢٤٦١٩

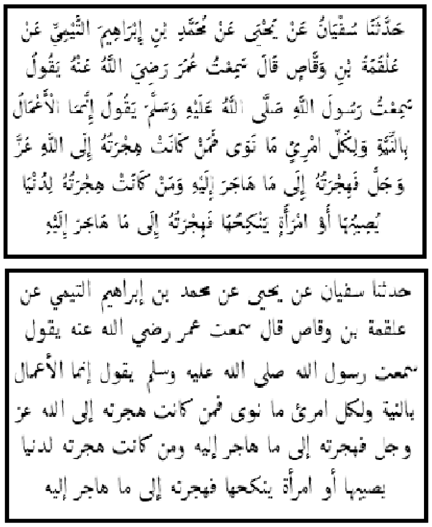
Statistics

Using Diacritics To Hide Data

Analysis indicates that in standard Arabic the frequency of one diacritic, namely Fatha, is almost equal to the occurrence of the other seven diacritics.

Assign a 1 to the diacritic Fatha and the remaining seven diacritics will represent a 0.

Use a cover media that is empty of diacritics.



حَدَّثَنَا سُفْيَانُ عَنْ يَحْيَى عَنْ مُحَمَّدِ بْنِ إِبْرَاهِيمَ التَّيْمِيِّ عَنْ
عَلْقَمَةَ بْنِ وَقَّاصٍ قَالَ سَمِعْتُ عُمَرَ رَضِيَ اللَّهُ عَنْهُ يَقُولُ
سَمِعْتُ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ يَقُولُ إِنَّمَا الْأَعْمَالُ
بِالنِّيَّةِ وَلِكُلِّ امْرِئٍ مَا نَوَى فَمَنْ كَانَتْ هِجْرَتُهُ إِلَى اللَّهِ عَزَّ
وَجَلَّ فَهِجْرَتُهُ إِلَى مَا هَاجَرَ إِلَيْهِ وَمَنْ كَانَتْ هِجْرَتُهُ لِدُنْيَا
يُصِيبُهَا أَوْ امْرَأَةٍ يَنْكِحُهَا فَهِجْرَتُهُ إِلَى مَا هَاجَرَ إِلَيْهِ

حدثنا سفيان عن يحيى عن محمد بن إبراهيم التيمي عن
علقمة بن وقاص قال سمعت عمر رضي الله عنه يقول
سمعت رسول الله صلى الله عليه وسلم يقول إنما الأعمال
بالنية ولكل امرئ ما نوى فمن كانت هجرته إلى الله عز
وجل فهجرته إلى ما هاجر إليه ومن كانت هجرته لدنيا
يصيبها أو امرأة ينكحها فهجرته إلى ما هاجر إليه

Diactrized and non-diactrized text

Syntactically Correct

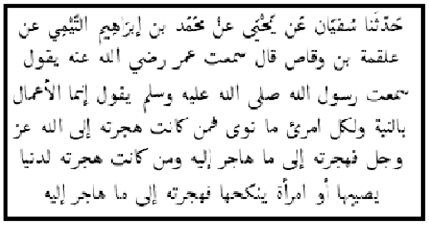
- To encode a value of 1 the algorithm looks for the first location where a Fatha can be placed and inserts the diacritic Fatha in the text.
- Location determination is based on the rules defined by the Standard Arabic language grammar and syntax.
- Or we can compare it to a copy of the cover media that is already diactrized (faster, and less complex)

Implementation Example

Next, the algorithm looks for the next location where a Fatha can be placed if another 1 needs to be inserted and adds the Fatha.

Otherwise, to insert a bit value of 0 the algorithm locates the first next position where any of the other diacritics can be inserted and adds that diacritic.

This process is repeated for as long as there are bits remaining to be hidden.



حَدَّثَنا سُفيَان عَن يَحيَى عن مُحَمَّد بن إبراهيم التَّيمِي عن
علقمة بن وقاص قال سمعت عمر رضي الله عنه يقول
سمعت رسول الله صلى الله عليه وسلم يقول إنما الأعمال
بالنية ولكل امرئ ما نوى فمن كانت هجرته إلى الله عز
وجل فهجرته إلى ما هاجر إليه ومن كانت هجرته لدنيا
يصيبها أو امرأة ينكحها فهجرته إلى ما هاجر إليه

Encoding the sequence 10101110101110000 using diacritics

قال الشيـخ الإمـام الـحـافظ أبـو عـبـد الـلـه
محمد بن إسـماعيـل بن إبراهـيـم بن
الـمـغيرة الـبخاري رحـمـه الـلـه تعـالـى
آمـين

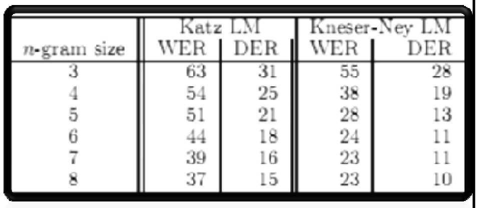
Encoding the same sequence using Kashida

Reusing The Cover Media

The output file will have less diacritics than the original cover media (because of deletion).

This means that reusing the same document more than once will mean less capacity.

A research group at IBM has proposed techniques for restoration of Arabic diacritics based on maximum entropy.



n-gram size	Katz LM		Kneser-Ney LM	
	WER	DER	WER	DER
3	63	31	55	28
4	54	25	38	19
5	51	21	28	13
6	44	18	24	11
7	39	16	23	11
8	37	15	23	10

Fig. 11: Error rate in % for n-gram diacritic restoration

Results

Compared to other techniques, capacity is the highest if a fully diactrized document is used as cover media.

Ambiguity is dependent on the reader's familiarity with Arabic language.

Robustness is high since it can withstand:

- Printing
- Retyping
- Font changing
- OCR

Table 1: Diacritics Technique

File Type	**File Size (Bytes)**	**Cover Size (Bytes)**	**Capacity (%)**
.txt	10,356	318,632	3.250 %
.wav	43,468	1,334,865	3.256 %
.jpg	23,796	717,135	3.318 %
.cpp	10,356	318,216	3.254 %
		Average	3.27 %

Table 2: Kashida Technique

File Type	**File Size (Bytes)**	**Cover Size (Bytes)**	**Capacity (%)**
.txt	4439	365181	1.215 %
.html	4439	378589	1.172 %
.cpp	10127	799577	1.266 %
.gif	188	15112	1.244 %
		Average	1.22 %

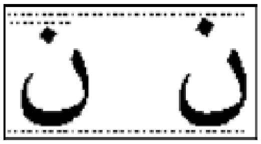
Analysis

Advantages	Disadvantages
Approach is easily implemented using software.	Medium to low ambiguity.
It produces high capacity.	Sending Arabic message with diacritics might raise suspicions nowadays.
Can be modified for more ambiguity (Use one of the diacritics as dummy diacritic, or as a switching diacritic)	Arabic font has different encodings on different machines, can be computer dependant.
Fairly robust. Can withstand OCR, retyping, printing and font changing.	

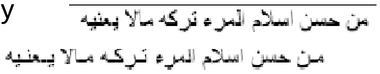
Using Multiple Diacritics in Arabic Scripts for Steganography

Remind: Related Arabic Stego-Work

- Shirali-Shahreza:
  - The position of dots is changed to render robust, yet hidden, information. The method needs special fonts.



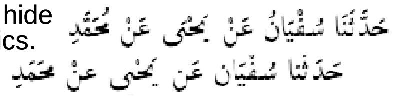
- Gutub:
  - Secret-bit hiding after dotted letters by inserting Kashidah's. A small drop in capacity occurs due to restriction of script on Kashidah and due to the extra-Kashidahs.



من حسن اسلام المرء تركه مالا يعنيه

مـن حسن اسلام المرء تـركـه مـالا يـعنيه

- Aabed et al.:
  - Redundancy in diacritics is used to hide information by omitting some diacritics.

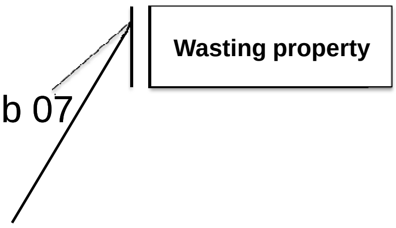


حَدَّثَنَا سُفْيَانُ عَنْ يَحْيَى عَنْ مُحَمَّدٍ

حَدَّثنا سُفْيَان عَن يَحْيَى عن مُحَمَّدٍ

Related Work

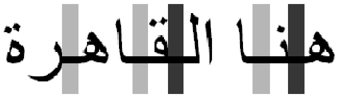
- Dots → Shahreza et. al 06
- Connectivity → Gutub 07
- Diacritics → Aabed et. al 07, Elarian et. al 07



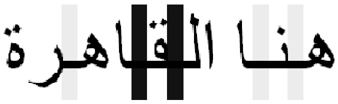
Wasting property

Wasting property

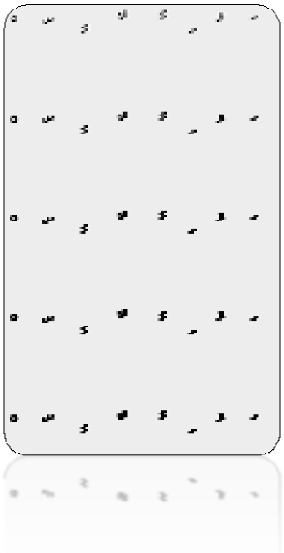
- *Wasting property*: different kinds of cover hiders for different types of secret patterns to be hidden.
  - Algorithm wasting secret bit *11001* carriers in COVER.



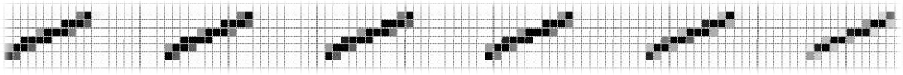
  - waste of cover hiders whenever the opposite-to-needed secret-bit holder is found before the needed one



Diacritics intensity



- Notice the differences in levels of darkness (as to the right) or colors (as below) in the single and the repeated diacritics.



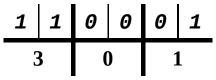
Hiding Scenario (1)

- 1st scenario (*Secret = 110001*)
  - – Direct *(block size = inf.)*

```
For each block $b_i = n_d$
    Repeat the $i^{th}$ diacritic $n_d$ times.
    Repeat the $1^{st}$ diacritic 49 extra times
```

Hiding Scenario (2)

- Blocked
- Block size=2



- – For *Secret = 110001*

Divide Secret into block of 2-bits

Repeat the first diacritic 3 times (3 = $(11)_b$),

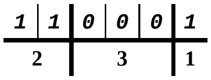
the second one 0 times (0 = $(00)_b$),

and the third one 1 time (1 = $(01)_b$).

Hiding Scenario (3)

- RLE (run-length enoding)

```
While(secret!=EOF & cover!=EOF
   b = b^
   While(secret.b = b)
         Type diacritic
```



  – For *Secret = 110001*

  Repeat the 1st diacritic 2 times (1’s in (*11*)$_b$); the 2nd one, 3 times (0’s in (*000*)$_b$); and the 3rd one, 1 time (for *1*).

Summary of the three scenarios

The encodings of the binary value 110001 according to the scenarios of the first approach

Scenario	Approach	Extra diacritics
1st scenario (stream)	Repeat the first diacritic 49 times. (49 = (*110001*)$_b$).	49.
2nd scenario block size=2	Repeat the first diacritic 3 times (3 = (*11*)$_b$), the second one 0 times (0 = (*00*)$_b$), and the third one 1 time (1 = (*01*)$_b$).	3 + 0 + 1 = 4.
3rd scenario (RLEstart=1)	Repeat the first diacritic 2 times (2 = number of 1’s in (*11*)$_b$), the second one 3 times (3 = number of 0’s in (*000*)$_b$), and the third one 1 time (for *1*).	(2-1) + (3-1) + (1-1) = 3.

Mapping the Hidden Message

Study Example: scenario 2 & 3

- The fixed-size scenario parses a stream of binary bits into blocks of fixed-size.

1	1	0	0	0	1	0	1
3		0		1		1	

- The variable-size content-based scenario parses a stream of binary data based on runs, regardless of the number of bits they occupy.

1	1	0	0	0	1	0	1
2		3			1	1	1

contd..

Study Example: scenario 2 & 3

Encodings of the binary value 11000101 according to two scenarios.

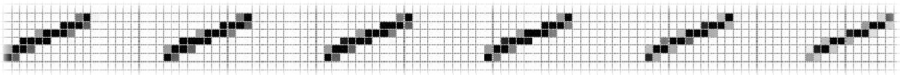
Scenario	Mapping	Needed Diacritics	Extra Diacritics
Fixed-size = 1	1 1 0 0 0 1 0 1 1 1 0 0 0 1 0 1	8	1 + 1 + 0 + 0 + 0 + 1 + 0 + 1 = 4
Fixed-size = 2	1 1 0 0 0 1 0 1 3 0 1 1	4	3 + 0 + 1 + 1 = 5
Fixed-size = 4	1 1 0 0 0 1 0 1 12 5	2	12 + 5 =13
Content-based	1 1 0 0 0 1 0 1 2 3 1 1 1	5	(2-1)+(3-1)+(1-1)+(1-1)+(1-1) = 8 - 5 =3

The textual approach

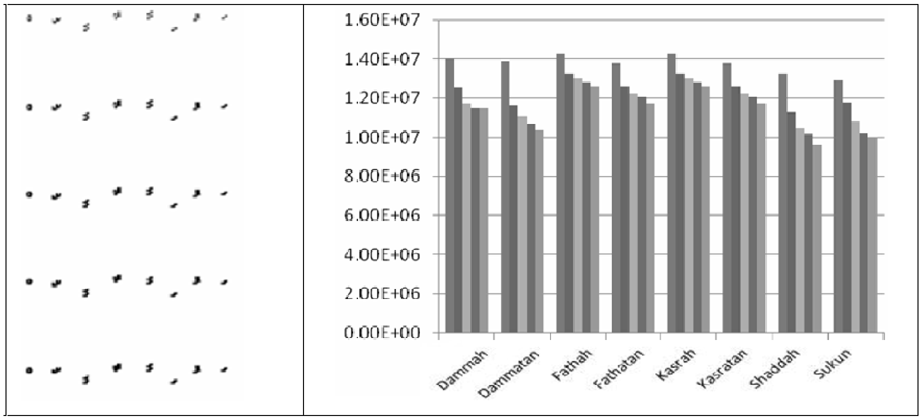
- The textual approach chooses a font that hides extra (or maybe all) diacritic marks completely.
- It uses any encoding scenario to hide secret bits in an arbitrary number of repeated but invisible diacritics.
- A softcopy of the file is needed to retrieve the hidden information (by special software or simply by changing the font).

The image approach,

- The image approach selects one of the fonts that slightly darken multiple occurrences of diacritics.
- This approach needs to convert the document into image form to survive printing.
- This unfortunate fact reduces the possible number of repetition of a diacritic to the one that can survive a printing and scanning process (up to 4 as the last two columns of the first diacritic



Approaches cont..



The degree of brightness of the diacritic marks repeated 1, 2, 3, 4 and 5 times each

---

Approaches cont..

Comparison between the two approaches in terms of capacity, robustness and security.

Approach	**Capacity**	**robustness**	**security**
Text + softcopy	High, up to infinity in 1st scenario	Not robust to printing	Invisible, but in the code
Image + softcopy	Very low, due to image overhead	Robust to printing	Slightly visible. Sizeable
Image + hardcopy	Moderate, 1st scenario, blocks of 2	Robust to printing	Slightly visible

Evaluation

The ratios of usable characters for hiding both binary levels according to the three studied approaches

Approach	***p***	***q***	***r***	**(*p*+*r*+*q*)/2**
Dots	0.2764	0.4313	0.0300	0.3689
Kashidah-Before	0.2757	0.4296	0.0298	0.3676
Kashidah-After	0.1880	0.2204	0.0028	0.2056
Diacritics	0.3633	0.3633	0	0.3633

Comparison
Diacritics *vs*. Kashidah

- Pros:
  - While Kashidah suffers from restrictions on its insertion, almost every character can bare a diacritic on it. This disadvantage of the Kashidah method becomes severer for the need of dotted character.
- Cons:
  - Diacritics never come alone; but with another character → a stable overhead of 2 bytes per secret-baring position.

Comparison Cntd.
The Advantage

- The advantage of our work is that each usable character can bare multiple secret bits with 1 character as overhead. Although this same overhead can be claimed in the Kashidah method, it can’t really be applied for Kashidah becomes too long and noticeable.

Diacritics Remarks

- The text and image approaches are discussed which are used to hide information in Arabic diacritics for steganographic.
- This work presents a variety of scenarios that may achieve up to arbitrary capacities. Sometimes tradeoffs between capacity, security and robustness imply that a particular scenario should be chosen.
- The overhead of using diacritics was, experimentally, shown very comparable to related works.
- The advantage of the method is that such overhead decreases if more than one diacritical secret bit is used at once.

Summary

- Steganography has its place in the security. On its own, it won’t serve much but when used as a layer of cryptography, it would lead to a greater security.
- Far fetched applications in privacy protection and intellectual property rights protection.
- Research is going on in both the directions
  - One is how to incorporate hidden or visible copyright information in various media, which would be published.
  - At the same time, in opposite direction, researcher are working on how to detect the trafficking of illicit material & covert messages published by certain outlawed groups.

- Too many images, MP3’s, spams, pictures, texts on the Internet.
- Too many possible algorithms.
- In conlusion:

too many secrets
setec astronomy