AN OVERVIEW ON HIDING AND DETECTING STEGO-DATA IN VIDEO STREAMS

Alexandre Miguel Ferreira

May 11, 2015

University of Amsterdam

Agenda

Research Question

Background

Literature Study

Analysis

Conclusion

RESEARCH QUESTION

Research Question

Which methods are available for (real-time) steganalysis on a video-stream and how can these be prevented?

- Which are the steganography methods available for video-stream?
- Which are the steganalysis methods available for video-stream?
- How can steganography be prevented on a video-stream?

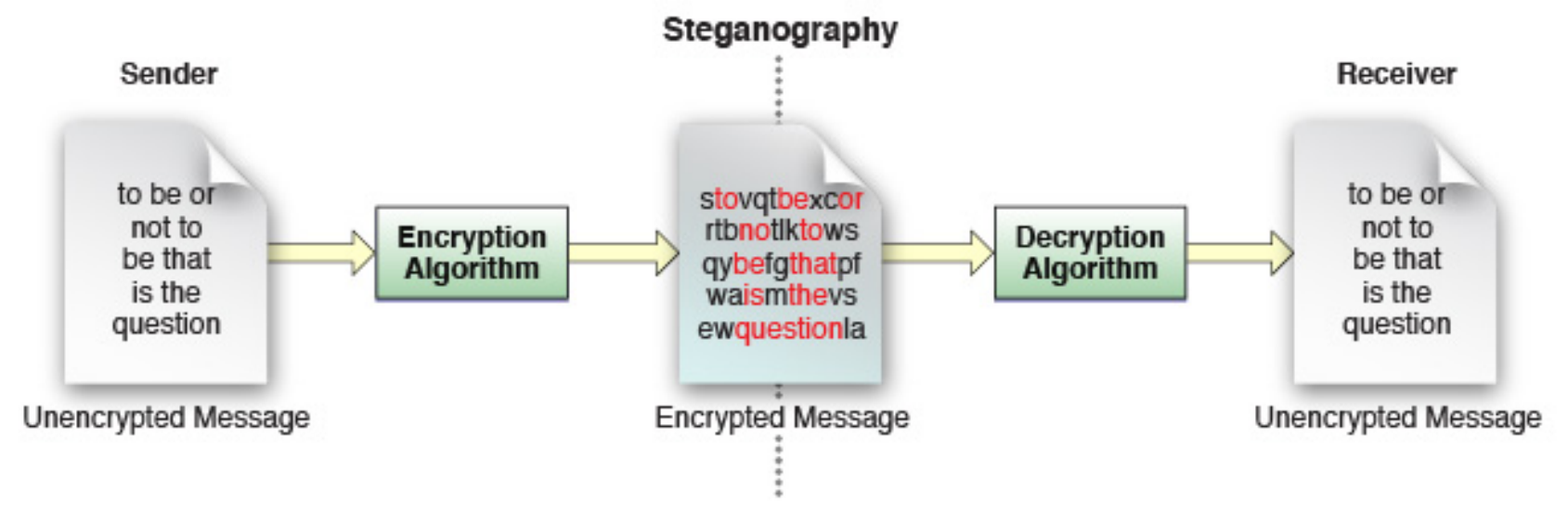
BACKGROUND

What is Steganography?

The art and science of hiding communication

Originates from the ancient Greek

- *steganos* (covered)
- *graphein* (writing)



Source: *https://developer.apple.com/*

What is Steganography? History

Earliest recordings from the Greek historian Herodotus (440 BC)

- Prisoners scalp tattooed to deliver secret messages
- Wooden tables carved before applying its wax surface

On the XV century Johannes Trithemius wrote about

- Invisible inks, Coding techniques for text, Hidden messages in music

Used to send hidden messages during World War II

- Null ciphers, Image substitution, Microdots

Steganography vs Watermarking

Similar to Steganography

- On Steganography the data embedded should be covert and undetectable
- On Watermarking it does not matter, however ...
- ... any attempt to remove it should result in significant degradation of the quality of the carrier file

Commonly used to help trace the origin of files

Steganography vs Cryptography

Different from Steganography

- Cryptography scrambles a message so it cannot be understood
- Steganography hides the message so it cannot be seen

Both are used to protect confidential information ...

- ... therefore often confused

What is Steganalysis?

Security of a steganographic system is defined by its strength to defeat detection

Practice of detecting the presence of messages that have been hidden using steganography

Ideally the content of the hidden message is also determined

What is Steganalysis? Types of Attacks

Steganalysis attacks can be active or passive

- On active attacks a steganalyst can manipulate the data
- On passive attack the steganalyst is only able to analyze the information without changing it

Attacks used by steganalysts to detect steganography on files can be:

- Visual Attacks
- Structural Attacks
- Statistical Attacks

Types of Attacks - Visual Attacks

The simplest form of attacking a steganographic system

Based on the visual analysis of the image

- Noticeable differences indicate that the image probably carries hidden information

If the carrier is not known this attacks becomes very hard

Types of Attacks - Structural Attacks

Analysis of known properties of the algorithms used to hide information

- Analysed further if found any properties of these algorithms

Outputs a lot of false positives

- Used to highlight images which show signs of possible embedding

Depends a lot on if the carrier file is known

Types of Attacks - Statistical Attacks

Statistical analysis done using mathematical formulas

- Much more effective than the Visual or Structural attacks

It is successful even without knowing the carrier file ...

- ... however it fails to determine the hidden data's size

LITERATURE STUDY

Steganographic Techniques (1)

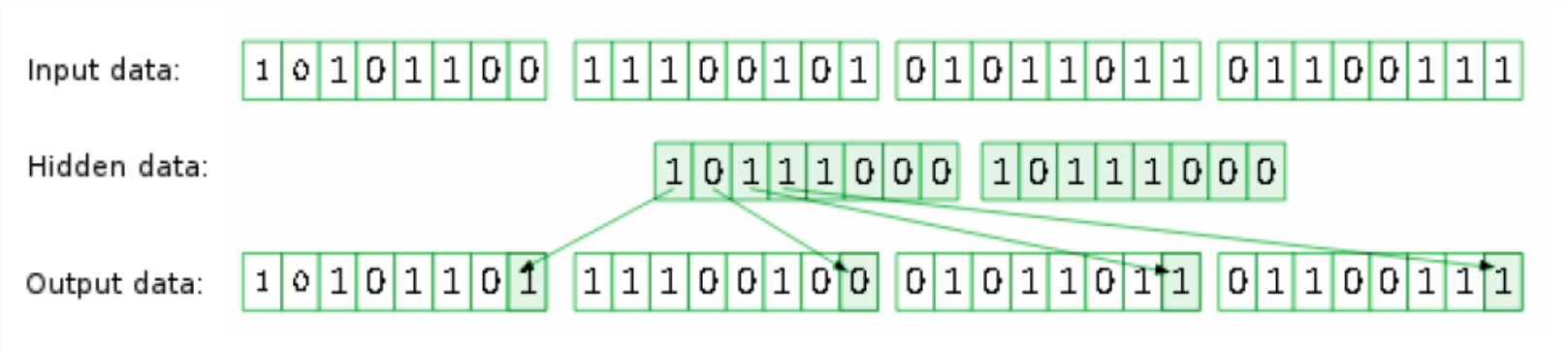
Big variety of techniques used to camouflage information:

- Injection
  - By far the simplest steganographic technique
  - Hides a message in parts of a file that are “ignored” by the application
- Substitution
  - Identify areas of a file of least relevance
  - Replace this data with the hidden information
  - Does not modify the size of the container file …
    - … therefore the steganographic capacity of the file is limited

Steganographic Techniques (2)

List Significant Bits Manipulation

- LSB Sequential Insertion
- LSB Pseudo Random Insertion
  - Pseudo Random Number Generator (PRNG) is used to randomly hide the secret bits of the message into the LSB of the carrier file



Source: *http://lvee.org/uploads/abstract_file/file/111/2.png*

Transform Domain

Generally used on compressed container files, such as JPEG or MPEG

- Discrete Cosine Transform
  - Algorithm works by using quantization
    - Rounding values of least important parts (not noticeable by the human eye)
  - Image is split into smaller areas to be transformed via DCT
    - Quantization on the frequencies is then applied
    - This is the stage where the secret message is injected
  - Finally the image is compressed
    - No impact on the integrity of the secret message
- Discrete Wavelet Transform
  - Makes it possible to rise the level of robustness of the information being hidden
  - If the threshold is too high the stego-file has detectable differences

Compression

Regards reducing and removing redundant video data ...

- ... with no undesirable effects on the visual quality

Lossless Compression

- Every single bit of data that was originally in the file remains after the file is uncompressed

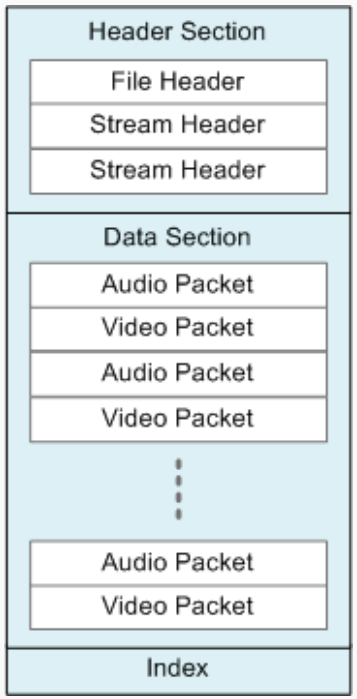
Lossy Compression

- Discards the points which are difficult to identify by the human eye
- Resulting image is similar to the original image
- Generally used on video and sound

Video Container Format

Contains the various components of a video

- Such as the stream of images or the sound



Source: *https://msdn.microsoft.com/*

ANALYSIS

Approach

Create some stego-videos

- *OppenPuff*

Perform known attacks

- Visual Attack
- Statistical Attack
- Structural Attack

OpenPuff (1)

Created by Cosimo Oliboni

The users to hide information in a wide range of carrier formats

- 3gp, Mp4, Mpeg II, etc.

Possible to hide data in more than a single carrier file

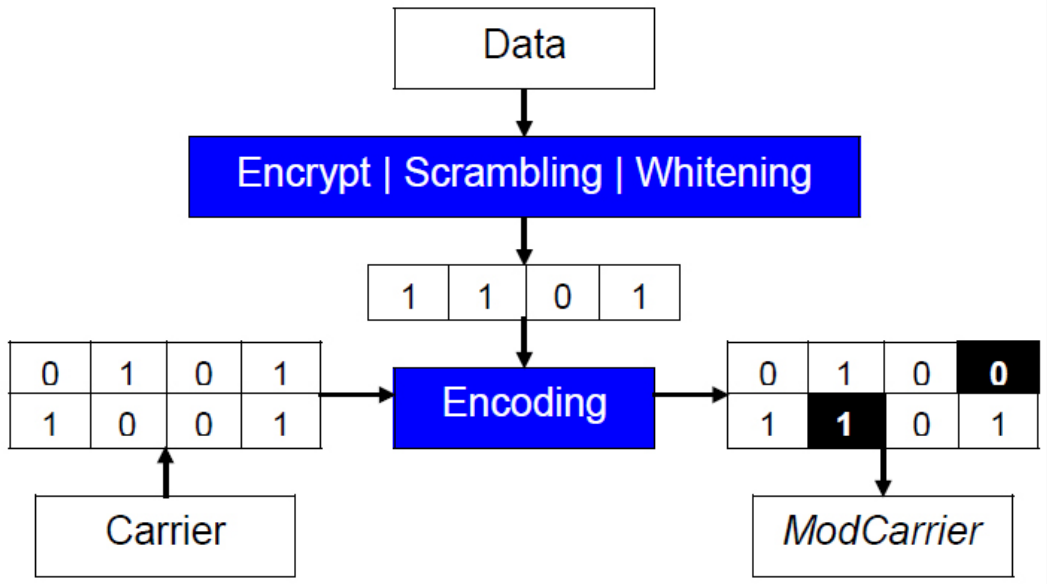
2 important factors were taken into consideration

- Embedding efficiency
- Embedding payload

OpenPuff (2)

Based on Niels Provos paper *Defending Against Statistical Steganalysis*

- which states ”steganalysis resistance and performance are incompatible trade-offs”



Source: *https://en.wikipedia.org/wiki/File:OpenPuff*

OpenPuff Stego-analyzed - Visual Attack

Performed by

- Reproducing both the original and stego videos
- Comparing and analysing individual frames from the original and from the stego-file

Original file frame

Stego-file frame

OpenPuff Stego-analyzed - Statistical Attack (1)

Program *ent* used to perform this attack

- **Entropy** - Information density of the contents of the file
- **Chi-square Test**
  - **greater than 99% and less than 1%** - almost surely not random
  - **between 99% and 95% or between 1% and 5%** - considered suspect
  - **between 90% and 95% or between 5% and 10%** - not sure to be suspect
- **Arithmetic Mean** - Result of the sum of all the bytes in the file divided by the its length
- **Monte Carlo Value for Pi** - If the sequence is close to random, the value will approach the correct value of π
- **Serial Correlation Coefficient** - Calculates how much each byte in the file depends on the previous byte

OpenPuff Stego-analyzed - Statistical Attack (2)

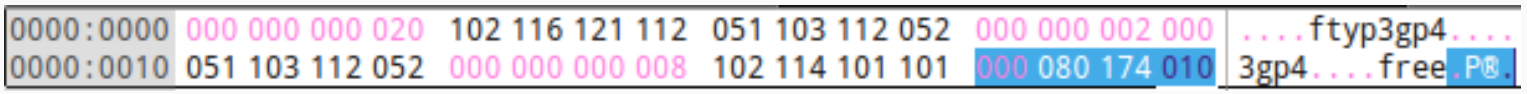
Values are very similar and do not raise any suspicious upon the stego-file

	Original	Stego	Expected
Entropy	1%	1%	0%
Chi-square Test	0.01%	0.01%	N/A
Arithmetic Mean	127.0006	126.5138	127.5
Monte Carlo Value for Pi	3.025822076	3.010476826	π
Serial Correlation Coefficient	0.147440	0.154106	0.0

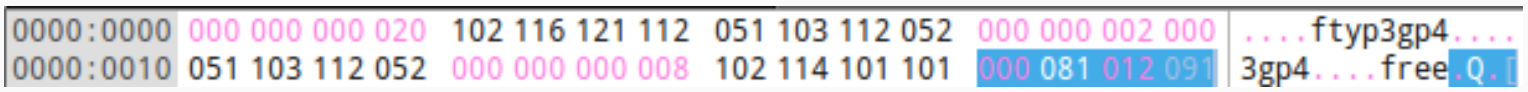
OpenPuff Stego-analyzed - Structural Attack (1)

Based on the comparison of the original file and the stego-file

- hexdump of both files was analyzed



File type header hexdump from the original file



File type header hexdump from the stego-file

OpenPuff Stego-analyzed - Structural Attack (2)

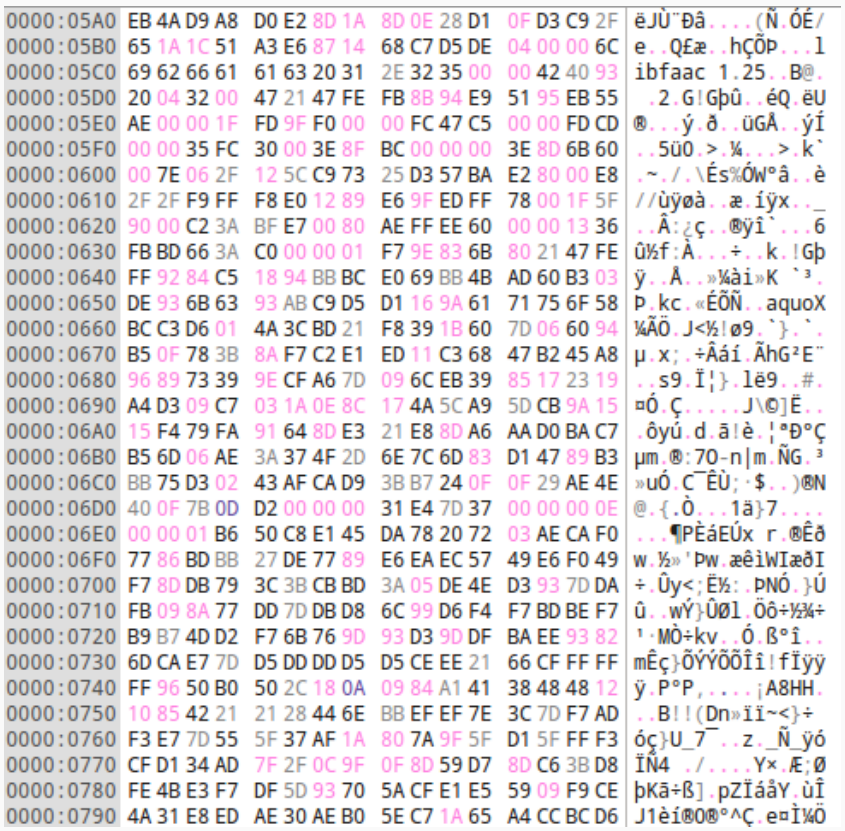
Last four bytes of the header are changed

- These bytes are an offset pointing to the beginning of the header that belongs to the MOOV box ...
- ... which defines the timescale, duration, display characteristics of the movie, as well as sub-boxes containing information for each track in the movie

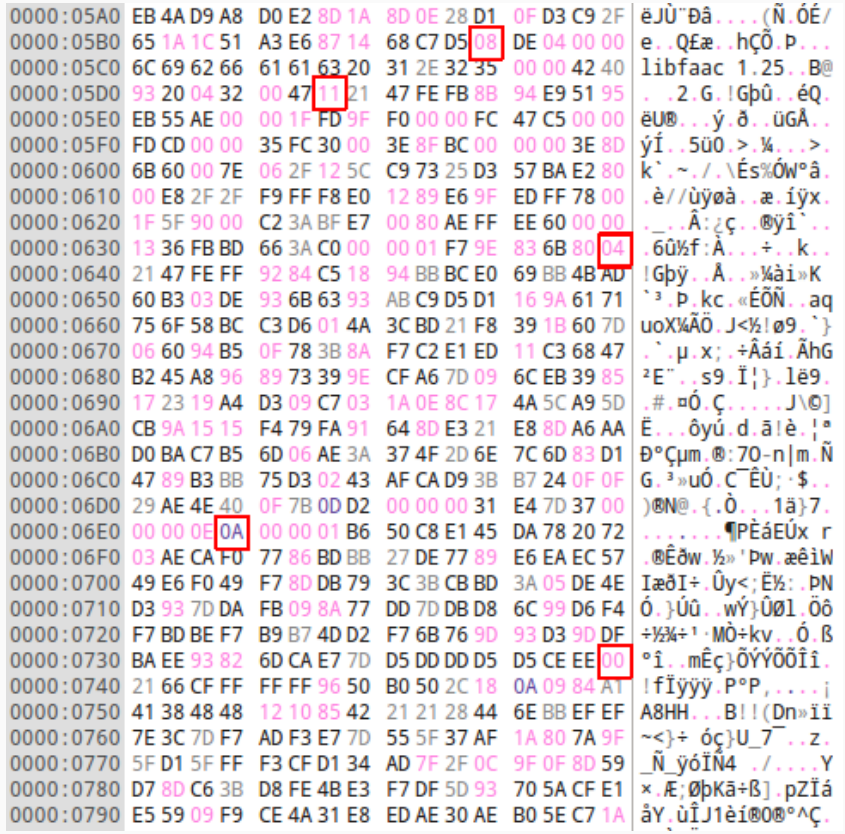
hexdump of both files is different since some bytes were inserted outside this box

OpenPuff Stego-analyzed - Structural Attack (3)

Pattern followed through out the stego-file, outside the MOOV box



Original file hexdump



Stego-file hexdump

OpenPuff Stego-analyzed - Structural Attack (4)

Although it could not be proved ...

- ... these bytes might be related to the size of the file being hidden
- ... as well as the password(s) used to encrypt the message

Assumption is made based on Niels Provos paper

- Stated that ”32 state bits are hidden, 16 bits for a seed and 16 bits for an integer containing the length of the message being hidden”

Important to notice that the video container format may change, therefore the optimal location of the moov box will be depend on this

OpenPuff Stego-analyzed - Structural Attack (5)

While analysing in detail the MOOV box, it was noticed that the bytes were modified

```
0003:8410 8C 73 74 63 6F 00 00 00  00 00 00 01 1F 00 00 00  .stco...........
0003:8420 24 00 00 06 E0 00 00 09  6B 00 00 14 59 00 00 1E  $...à...k...Y...
0003:8430 F5 00 00 22 7A 00 00 25  6A 00 00 28 71 00 00 2A  õ..."z..%j..(q..*
0003:8440 BD 00 00 2E 02 00 00 30  AF 00 00 35 33 00 00 3C  ½.......0¯..53..<
0003:8450 D4 00 00 44 A6 00 00 48  29 00 00 4B 15 00 00 4D  Ô..D¦..H)..K...M
0003:8460 8B 00 00 50 7D 00 00 53  B5 00 00 56 C2 00 00 59  ...P}..Sµ..VÂ..Y
0003:8470 78 00 00 5C 9A 00 00 5F  56 00 00 61 EF 00 00 63  x..\..._V..aï..c
0003:8480 8B 00 00 65 F2 00 00 68  65 00 00 6A D2 00 00 6D  ...eò..he..jÒ..m
0003:8490 48 00 00 70 29 00 00 74  C7 00 00 79 94 00 00 7D  H..p)..tÇ..y...}
0003:84A0 15 00 00 80 10 00 00 83  80 00 00 86 95 00 00 89  ................
0003:84B0 9F 00 00 8C 2F 00 00 8E  D9 00 00 91 4E 00 00 92  ..../...Ù...N...
0003:84C0 CB 00 00 94 D1 00 00 96  EF 00 00 99 12 00 00 9F  Ë...Ñ...ï.......
0003:84D0 9A 00 00 A6 40 00 00 A9  E8 00 00 AC 79 00 00 AE  ...¦@...©è..¬y..®
0003:84E0 A8 00 00 B1 F9 00 00 B4  CA 00 00 B7 D1 00 00 BA  ¨..±ù..´Ê..·Ñ..º
0003:84F0 66 00 00 BC ED 00 00 BF  38 00 00 C3 C7 00 00 C8  f..¼í..¿8..ÃÇ..È
0003:8500 96 00 00 CB 46 00 00 CE  92 00 00 D1 D4 00 00 D4  ...ËF..Î...ÑÔ..Ô
0003:8510 DC 00 00 D7 C3 00 00 DA  93 00 00 DD 84 00 00 DF  Ü..×Ã..Ú...Ý...ß
0003:8520 83 00 00 E1 D6 00 00 E8  BF 00 00 EF B3 00 00 F3  ...áÖ..è¿..ï³..ó
0003:8530 16 00 00 F6 13 00 00 F9  07 00 00 FB D4 00 00 FD  ...ö...ù...ûÔ..ý
0003:8540 A7 00 01 00 6C 00 01 02  E6 00 01 05 4C 00 01 07  §...l...æ...L...
0003:8550 A9 00 01 09 F5 00 01 0C  14 00 01 0E 6A 00 01 0F  ©...õ.......j...
0003:8560 C0 00 01 11 DB 00 01 13  EA 00 01 16 50 00 01 1A  À...Û...ê...P...
0003:8570 55 00 01 1D D7 00 01 21  21 00 01 23 C2 00 01 25  U...×..!!..#Â..%
0003:8580 55 00 01 27 DC 00 01 2A  31 00 01 2C A1 00 01 2E  U..'Ü..*1..,¡...
0003:8590 E4 00 01 31 1E 00 01 33  B6 00 01 35 FE 00 01 37  ä..1...3¶..5þ..7
0003:85A0 B8 00 01 3A 2D 00 01 3C  A4 00 01 3F 27 00 01 41  ¸..:-..<¤..?'..A
0003:85B0 69 00 01 4C C8 00 01 4F  CE 00 01 53 24 00 01 55  i..LÈ..OÎ..S$..U
0003:85C0 90 00 01 58 A1 00 01 5B  88 00 01 5D DE 00 01 64  ...X¡..[...]Þ..d
0003:85D0 06 00 01 6A 17 00 01 6C  CA 00 01 6F 6F 00 01 71  ...j...lÊ..oo..q
```

Original file MOOV box hexdump

```
0003:8860 04 8C 73 74  63 6F 00 11  04 20 00 00  01 1F 00 00  ..stco... ......
0003:8870 00 24 00 00  06 E4 00 00  09 73 00 00  14 65 00 00  .$...ä...s...e..
0003:8880 1F 05 00 00  22 8E 00 00  25 82 00 00  28 8D 00 00  ....".....%...(...
0003:8890 2A DC 00 00  2E 25 00 00  30 D6 00 00  35 5E 00 00  *Ü...%..0Ö..5^..
0003:88A0 3D 03 00 00  44 D9 00 00  48 60 00 00  4B 50 00 00  =...DÙ..H`..KP..
0003:88B0 4D C9 00 00  50 BF 00 00  53 FB 00 00  57 0C 00 00  MÉ..P¿..Sû..W...
0003:88C0 59 C6 00 00  5C EC 00 00  5F AC 00 00  62 49 00 00  YÆ..\ì.._¬..bI..
0003:88D0 63 E8 00 00  66 53 00 00  68 CA 00 00  6B 3B 00 00  cè..fS..hÊ..k;..
0003:88E0 6D B5 00 00  70 9A 00 00  75 3C 00 00  7A 0D 00 00  mµ..p...u<..z...
0003:88F0 7D 91 00 00  80 90 00 00  84 04 00 00  87 1D 00 00  }...............
0003:8900 8A 2B 00 00  8C BF 00 00  8F 6D 00 00  91 E6 00 00  .+...¿...m...æ..
0003:8910 93 66 00 00  95 70 00 00  97 92 00 00  99 B9 00 00  .f...p.......¹..
0003:8920 A0 45 00 00  A6 EF 00 00  AA 9B 00 00  AD 30 00 00  .E..¦ï..ª....0..
0003:8930 AF 62 00 00  B2 B7 00 00  B5 8C 00 00  B8 97 00 00  ¯b..²·..µ.....
0003:8940 BB 30 00 00  BD BB 00 00  C0 0A 00 00  C4 9D 00 00  »0..½»..À...Ä...
0003:8950 C9 6F 00 00  CC 23 00 00  CF 73 00 00  D2 B9 00 00  Éo..Ì#..Ïs..Ò¹..
0003:8960 D5 C5 00 00  D8 B0 00 00  DB 84 00 00  DE 79 00 00  ÕÅ..Ø°..Û...Þy..
0003:8970 E0 7B 00 00  E2 D2 00 00  E9 BF 00 00  F0 B7 00 00  à{..âÒ..é¿..ð·..
0003:8980 F4 1E 00 00  F7 1F 00 00  FA 17 00 00  FC E8 00 00  ô...÷...ú...üè..
0003:8990 FE BE 00 01  01 87 00 01  04 05 00 01  06 6F 00 01  þ¾.............o..
0003:89A0 08 D0 00 01  0B 20 00 01  0D 43 00 01  0F 9D 00 01  .Ð... ...C......
0003:89B0 10 F6 00 01  13 15 00 01  15 28 00 01  17 92 00 01  .ö.......(......
0003:89C0 1B 9B 00 01  1F 21 00 01  22 6F 00 01  25 14 00 01  .....!.."o..%...
0003:89D0 26 AA 00 01  29 35 00 01  2B 8E 00 01  2E 02 00 01  &ª..)5..+.......
0003:89E0 30 49 00 01  32 87 00 01  35 23 00 01  37 6F 00 01  0I..2...5#..7o..
0003:89F0 39 2C 00 01  3B A5 00 01  3E 20 00 01  40 A7 00 01  9,..;¥..>  ..@§..
0003:8A00 42 ED 00 01  4E 50 00 01  51 5A 00 01  54 B4 00 01  Bí..NP..QZ..T´..
0003:8A10 57 23 00 01  5A 38 00 01  5D 23 00 01  5F 7D 00 01  W#..Z8..]#.._}..
0003:8A20 65 A9 00 01  6B BE 00 01  6E 75 00 01  71 1E 00 01  e©..k¾..nu..q...
```

Stego-file MOOV box hexdump

OpenPuff Stego-analyzed - Structural Attack (6)

Secret information is hidden inside the the MOOV box

Once again it could not be proved ...

... due to two reasons:

- The fact that the secret information is encrypted
- The use of deniable steganography techniques

Anti-Forensics

Pursuits to make the analysis and/or examination of evidence difficult or impossible to conduct

- Encryption and steganography among the ways

Relies on several weaknesses of the forensic process

- Human element, dependency on tools

There is always the chance of being detected using these techniques

- Resisting to these unpredictable attacks is also possible ...
- ... even when forced to provide a valid password to extract the data

Anti-Forensics - Deniable Steganography

Camouflage based technique

- Even if the steganalyst is able to state that data is being hidden, allows the breaker to convincingly deny that fact

OpenPuff implements deniable steganography

- Possible to hide two different messages in the cover file
  - One which contains the sensitive data
  - One which although is plausible to be considered sensitive, the user is willingly to give away

One of the reasons why the statistical attacks are ineffective

CONCLUSION

Conclusion

Techniques used on images and audio can also be applied to videos

- Most common use the spacial domain (LSB) and the frequency domain (DCT)

Statistical analysis can reveal the presence of hidden data

- However it is a difficult process to carry out
- Hidden information tends to be nearly impossible to be detectable

Best way to prevent steganography would be to alter or destroy files which are considered suspicious

- New video compression methods where less redundant bits are available is also a possibility

Future Work

The attacks performed proved to be insufficient to determine the hidden information

- It would be interesting to assess if the hidden information can be retrieved

Questions?