Steganography

Art and science of communicating in a way that hides the existence of a message

- signal or pattern imposed on content
  - Persistent under transmission
  - Not encryption – original image/file is intact
  - Not fingerprinting
    • Fingerprinting leaves separate file describing contents

Classic techniques

- Invisible ink (1st century AD - WW II)
- Tattooed message on head
- Overwrite select characters in printed type in pencil
  - look for the gloss
- Pin punctures in type
- Microdots (early 20th century)
- Newspaper clippings, knitting instructions, XOXO signatures, report cards, …

Motivation

- Steganography received little attention in computing
- Renewed interest because of industry’s desire to protect copyrighted digital work
  - Audio, images, video, documents
- Detect counterfeiter, unauthorized presentation, embed key, embed author ID
- Also useful for forensics: enemies may use steganography to conceal their messages
  - Communication, stolen data, botnet controls

Steganography ≠ Copy protection

Isis and al-Qaeda sending coded messages through eBay, pornography and Reddit

Kashmira Gander – Monday 2 March 2015 19:29 GMT

Isis and al-Qaeda members are communicating with each other via coded messages hidden on websites including eBay, Reddit, and inside pornographic photos, according to a new book.

Gordon Thomas, who has sources inside Israel’s Mossad spy agency, has revealed that the organisation’s cyber warfare department’s most skilled cryptologists mastered a technique known as steganography, which is used to conceal secret information within a digital file. The spies found that al-Qaeda had used the technique to hide messages in goods offered for sale on eBay, according to extracts from Gabriel’s Spies: The Secret History of the Mossad published by The New York Post.
Null Cipher

Hide message among irrelevant data
Confuse the cryptoanalyst

Big rumble in New Guinea.
The war on celebrity acts should end soon. Over four big ecstatic elephants replicated!

Bring two cases of beer.

Chaffing & Winnowing

- Separate good messages from the bad ones
  - Easy for someone who has the key, difficult for someone who does not
- Stream of un-encoded messages with signatures or MACs
  - Some signatures are bogus
  - Need to have the key to test

For example:

<table>
<thead>
<tr>
<th>Alice</th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>M3</td>
</tr>
<tr>
<td>M3</td>
<td>M0</td>
</tr>
<tr>
<td>M1</td>
<td>M2</td>
</tr>
<tr>
<td>M2</td>
<td>M1</td>
</tr>
</tbody>
</table>

Irene?

Steganography in images

Spatial domain
- Bit flipping
- Color separation

Frequency domain
- Embed signal in select frequency bands
  (e.g., high frequency areas)
- Apply FFT/DCT transform first
- Alter the least perceptible bits to avoid detection
  - But watch out: these are the same bits targeted by lossy image compression software (such as jpeg)

BBC News 27 April 2006

Judge creates own Da Vinci code

The judge who presided over the failed Da Vinci Code plagiarism case at London's High Court had his own secret code in his written judgement.

Seemingly random italicised letters were included in the 71-page judgement given by Mr Justice Peter Smith, which apparently spell out a message.

Mr Justice Smith said he would confirm the code if someone broke it.

"I can't discuss the judgement, but I don't see why a judgement should not be a matter of fun," he said.

Italicised letters in the first few pages spell out "Smithy Code", while the following pages also contain marked out letters.
There are differences – but you don’t notice them in the photo

**Video**

- Coding still frames - spatial or frequency
- Data encoded during refresh
  - closed captioning
- Visible watermarking
  - used by most networks (logo at bottom-right)

**Audio**

Perceptual coding
- Inject signal into areas that will not be detected by humans
- May be obliterated by compression

Amazon MP3 audio
Identifies where the song was purchased, not the user

**Machine ID codes in laser printers**

See http://www.eff.org/Privacy/printers/

Designed by Xerox to identify counterfeit currency and help track down counterfeiters
UV Watermarking

Also passports, hand stamps for amusement park re-entry.

Text

- Text lines shifted up/down
  (40 lines text $\Rightarrow 2^{40}$ codes)
- word space coding
- character encoding - minor changes to shapes of characters

more more

Text

- Text lines shifted up/down
  (40 lines text $\Rightarrow 2^{40}$ codes)
- word space coding
- character encoding - minor changes to shapes of characters

• works only on “images” of text e.g., PDF, postscript

Text-based steganography

“Apparently, during the 1980’s, British Prime Minister Margaret Thatcher became so irritated at press leaks of cabinet documents that she had the word processors programmed to encode their identity in the word spacing of documents, so that disloyal ministers could be traced.”

≈ Ross Anderson
Streching the Limits of Steganography

Watermarking vs. Steganography

Both techniques hide a message in data

Goal of steganography
- Intruder cannot detect the message
- Primarily 1:1 communication

Goal of watermarking
- Intruder cannot remove or replace the message (robustness is important)
- Doesn’t have to be invisible
- Primarily 1:many communication

Watermarking applications

- Copyright protection
  - Embed information about owner
- Copy protection
  - Embed rights management information
  - But you need a trusted player
- Content authentication
  - Detect changes to the content
Content Protection

Content protection
• Digital content is simple to copy and distribute
  – Software, music, video, documents
• That’s not always good
  – How do software companies & artists make a living if their content is freely distributed on a large scale?
  • Maintain revenue streams
  • Enforce distribution rights (e.g., video available in the U.S. first)
• How do organizations keep their documents secure?
  • Enforce confidentiality & protect trade secrets?
  
  How can we make illegal content access difficult?

DRM
• Content industry (movies, music, documents) asked for technical solutions to the content distribution problem
• This led to digital rights management (DRM)
  – Protection of content
  – Definition on how it can be played and copied

  Not just documents & movies:
  – Printer cartridges
  – John Deere tractors
  – Keurig coffeemakers
  • RFID connections enforce use of Keurig-branded K-cups

Copy or execution protection
• On-device or on-installation checks
  – Software is configured to check a computer ID or license key when run
  – May validate online via a server
• Continuous or periodic network checks
  – Software must contact an on-line license server & identify itself and the computer to run
  • Subscription services do this: Adobe, Autodesk, Microsoft
• Timebombs
  – Software ceases to function if it's found to be illegally installed
  – Illegal in some places

  All checks can be defeated
  Goal: balance technical difficulties, user convenience, and legal repercussions

Cloud software
Ultimate protection
• Company provides both the computing platform and the software
  • And you don’t have physical access to the platform
  • If your subscription expires, you cannot use the platform
Documents & Books

- Trusted readers & content management
  - Microsoft Office, Adobe PDF, AutoCAD

- E-book readers
  - EPUB (default format for Apple)
  - MOBI (MobiPocket, purchased by Amazon in 2005)
  - AZW, AZW3 (Amazon Kindle – similar to MOBI)
  - PDF (Adobe)

MOBI/AZE formats

- MobiPocket
  - Acquired by Amazon in 2005 and used in Kindle
  - Reverse engineered & source published

- Uses a PC1 symmetric cipher with a 128-bit key
  - key = encrypted with temp key
  - temp key = Encrypt device ID with global_secret_key

- Main weakness
  - Device ID is 7 alphanumeric chars, only upper-case, followed by $'

- Kindle AZW format
  - Essentially MobiPocket with a device ID that ends with * instead of $

Media

Apple FairPlay

- AAC & MP4 files
- Content encrypted via AES
- Encrypted master key stored in the MP4 container file
- User key decrypts master key
  - User obtains user key when device authorizes with Apple’s servers
- Has been reverse-engineered

\[
\text{user_key} \\
\text{E}_{\text{user_key}}(\text{master_key}) \quad \text{E}_{\text{master_key}}(\text{content})
\]

Microsoft Protected Environment (PE)

- Enable protected content to flow through the OS in a protected manner.
  - All components in the protected environment must be trusted.
  - Trusted component = components signed by Microsoft, including kernel modules
  - OS stops DRM-restricted content from playing while unsigned software is running

- Content flows through trusted components:
  - Protected Media Path (includes Protected Video Path)

- Media source specifies the rights for using the content
  - Play, transfer, etc.

- Final output
  - Decrypted, uncompressed video frames travel on a physical connector to the display device
  - Providers may require protection in this area, such as the use of High-Bandwidth Digital Content Protection (HDCP) or DisplayPort Content Protection (DPCP)

© 2019 Paul Krzyzanowski
Broadcast Video

- Content is encrypted
- Key is transmitted via **Entitlement Management Messages (EMM)**
- The trick is to send the key so that only allowed users can get it

Digital Video Broadcasting

- **Relies on trusted hardware**
  - Data stream is decrypted with smart cards containing subscriber info
- **Source content is encrypted with a 48-bit secret key** *(key = control word)*
  - Control word may change several times per minute
  - Control word is encrypted & sent to all subscribers as part of an **Entitlement Control Message (ECM)**
- **Key to decrypt the ECM is sent to each subscriber as an Entitlement Management Message (EMM)**
  - Sent at less frequent intervals (several days to several weeks)
  - Encrypted per subscriber for their smart card

CableCARD

- Card device to allow customers to access digital cable TV channels on generic devices
- Identifies and authorizes subscriber
- Receives EMM (Entitlement Management Messages) for premium channels
- Decodes encrypted digital cable signal
- Performs conditional access logic & decryption
- Provides an MPEG-2 media stream to the host
- Tuner and MPEG decoder are part of the host equipment

CableCARD did not provide **host device certification for two-way communication**
- Deployment of proprietary set-top boxes is far bigger than CableCard
- Next (possible) successor: **AllVid**
  - Universal adapter for all types of pay TV and interactive program guides
  - Can communicate to any device with a screen
  - Endorsed by Google, Best Buy, Mitsubishi, Sony Electronics, TiVo
  - Not endorsed by cable companies
### DVD Content Scrambling System (CSS)

- Stream cipher – weak – based on 25-bit key – can be broken in $2^{25}$ tries
- Each player has one or more manufacturer-specific keys
- Each DVD has a disk key encrypted under each of the manufacturer’s keys
  - Goal was to produce new disks that omit a specific manufacturer’s key if it leaked
  - BUT – given any key in the system, all others can be found
  - Manufacturers had an incentive to keep costs down, not use tamper-resistant hardware
- DVD players on PCs
  - PCs are an open platform – only way to “protect” the code was to obfuscate it

### Blu-ray: Advanced Access Control System

- Blu-ray disc contains
  - Encrypted content: encrypted with a Title Key
  - Encrypted Title Key: Encrypted with a Volume Unique Key (Kvu)
  - Volume ID (VID): serial number of disc – will not be duplicated
  - Media Key Block (MKB) = lots of encrypted keys (~50 GB)
  - Allows each compliant device, using its secret device key, to compute a Media Key

- Player contains
  - One or more secret Device Keys
  - 128-bit keys provided to trusted parties by the AACS org
  - Device Keys may be unique per device or – often – shared by multiple devices

- Decryption
  - Use Device Key to decrypt a Media Key (Km) from the Media Key Block (MKB)
  - Combine Media Key & Volume ID to get the Volume Unique Key (Kvu)
  - Use Volume Unique Key to decrypt the Encrypted Title Key
  - Decrypt the content with the Title Key

### Content isn’t really protected

People built databases of media keys – so no need to decrypt the media key
- Do a googling search for AACS KEYDB.cfg
- https://gist.github.com/HenkPoley/41ed899251aa771cb1d061d49a388ae5
- 18 processing keys
- 23,999 titles as of 9/3/2017

There’s also the analog hole

### Widevine Content Protection

- Google’s Widevine DRM Platform
  - Used by Google Play Movies, Hulu, Spotify, Netflix, Amazon Prime Video, Disney+
  - Supported by:
    - Chromium, Firefox, Opera (not Firefox or MS Edge)
    - Most smart TVs
  - Over 30 chipsets support Widevine
    - ARM Trusted Execution Environment (TEE) handles:
      - Rights management, integrity management, firmware updates, authentication
    - Google licenses the code only to approved developers, devices, and applications
      - This is closed source software!
Widevine security levels

**Level 1 (L1)**
- All content processing, cryptography & control must be performed within the TEE
- Usually needed to access HD content

**Level 2 (L2)**
- Cryptography must be performed within the TEE but not video processing

**Level 3 (L3)**
- Used when the device does not have a TEE or processing is done outside the TEE
- Appropriate measures must be taken to protect the media stream within the host OS
- Broken in Jan 2019

Legal barriers: DMCA

**Digital Millennium Copyright Act**
Criminalizes production and dissemination of technology, devices, or services intended to circumvent measures (DRM) that control access to copyrighted works. It also criminalizes the act of circumventing an access control, whether or not there is actual infringement of copyright itself.

Without DMCA, anyone would be able to build a set-top box to decode video signals
- Just crack HDCP (High Definition Content Protection)

Also
- Licensing agreements (EULAs)
- EU’s Copyright Directive

The end