Computer Security

14. Content Protection & Steganography

Paul Krzyzanowski

Rutgers University

Spring 2017

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?
  - How do organizations keep their documents secure?

How can we make distribution more difficult?

Associate software with a computer

- Find unique characteristics of a machine
  1. CPU serial number (early microprocessors didn't have these)
  2. Add a dongle (USB hardware key)
  3. Create a unique ID based on PC's configuration
  4. Install software in a way that cannot be copied
     (e.g., mark blocks as bad)

     Used on early PCs but not viable with modern operating systems

But

– You can go through the software with a debugger & remove checks

- This becomes harder as software gets bigger

Copy or execution protection

- On-device checks
  - Software is configured to check a computer ID or license key when run
- Network checks
  - Software must contact an on-line license server & identify itself and the computer to run
- 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

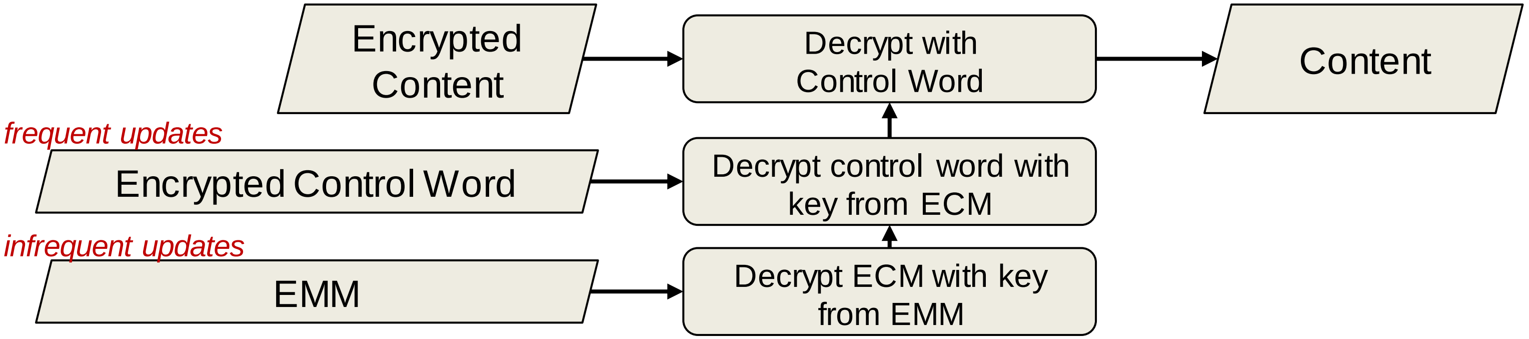
- Ultimate protection
  - Company provides computing platform and the software
  - If your subscription expires, you cannot use the platform

DRM

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

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)
- Ability 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


Encrypted Content
Decrypt with Control Word
Content
frequent updates
Encrypted Control Word
Decrypt control word with key from ECM
infrequent updates
EMM
Decrypt ECM with key from EMM

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

- CableCARD didn't 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 – 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 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

- Uses AES to encrypt content
  - Media key encrypted with a combination of a media key and volume ID (serial number of the disc)
    - Serial number cannot be duplicated on recordable media

- CSS
  - Unique encryption key for content – key is encrypted for a set of players
  - All players of a model group have the same decryption key
  - Disc contains several hundred encrypted keys, one for each licensed player model

- AACS
  - Unique media key for content – key is encrypted for a player
  - Each individual player has a unique set of decryption keys
  - Licensors can revoke keys for individual players in future content
  - AACS keys compromised since 2007 – keys were found using debuggers

Content isn't really protected

- People built databases of media keys – so no need to decrypt the media key
  - http://www.labdv.com/aacs/KEYDB.cfg
  - 18 processing keys
  - 20,822 titles as of 4/17/2017

There's also the analog hole

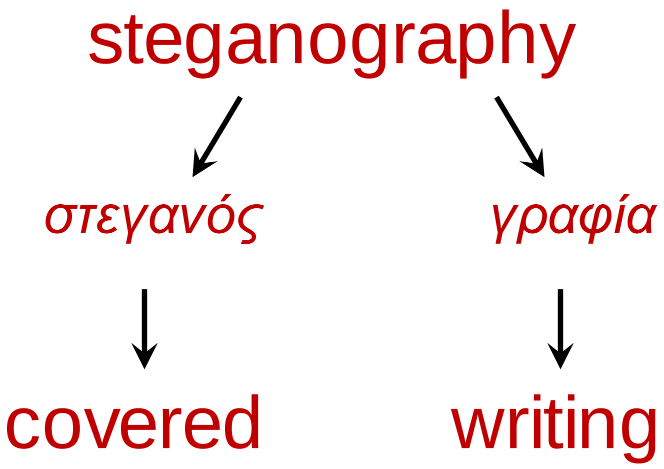
Legal barriers: DMCA

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)



steganography

στεγανός

γραφία

covered

writing

The art of secret (hidden) writing

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 (WW II)
- 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

INDEPENDENT

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 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 *Gideon's Spies: The Secret History of the Mossad* published by *The New York Post*.

Null Cipher

- Hide message among irrelevant data
- Confuse the cryptanalyst

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

Null Cipher

- Hide message among irrelevant data
- Confuse the cryptanalyst

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

Bring two cases of beer.

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 hid 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.

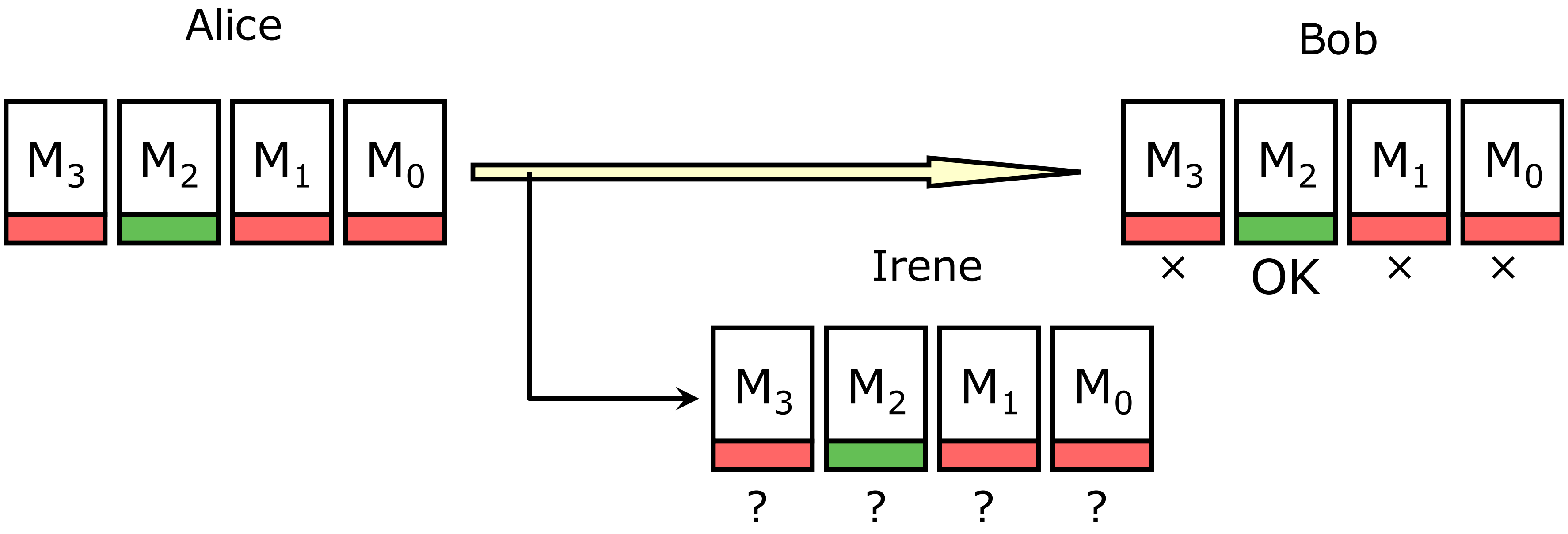
"I can't discuss the judgement, but I don't see why a judgement should not be a matter of fun," he said.

Italicised letters in the first few pages spell out "Smithy Code", while the following pages also contain marked out letters.

http://news.bbc.co.uk/go/pr/fr/-/1/hi/entertainment/4949488.stm

Chaffing & Winnowing

- Separate good messages from the bad ones
- Stream of unencoded messages with signatures
  - Some signatures are bogus
  - Need key to test

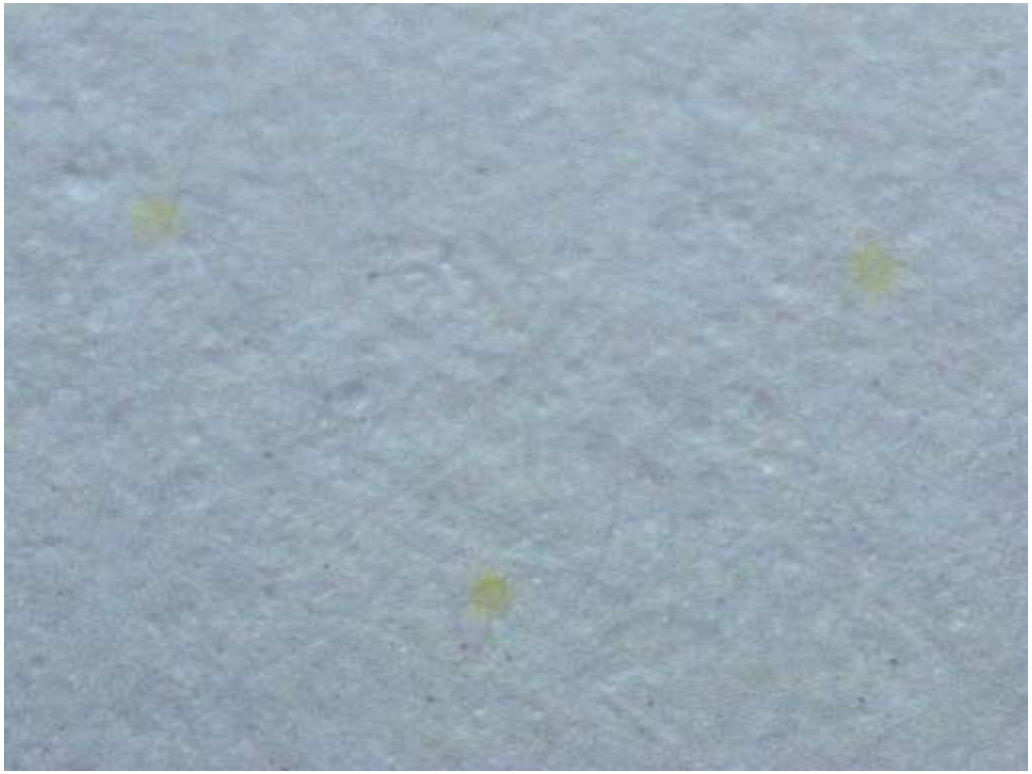


Steganography in images

- Spatial domain
  - bit flipping
  - color separation
- Frequency domain
  - embed signal in select frequency bands (e.g., high frequency areas)
  - apply FFT/DCT transform first

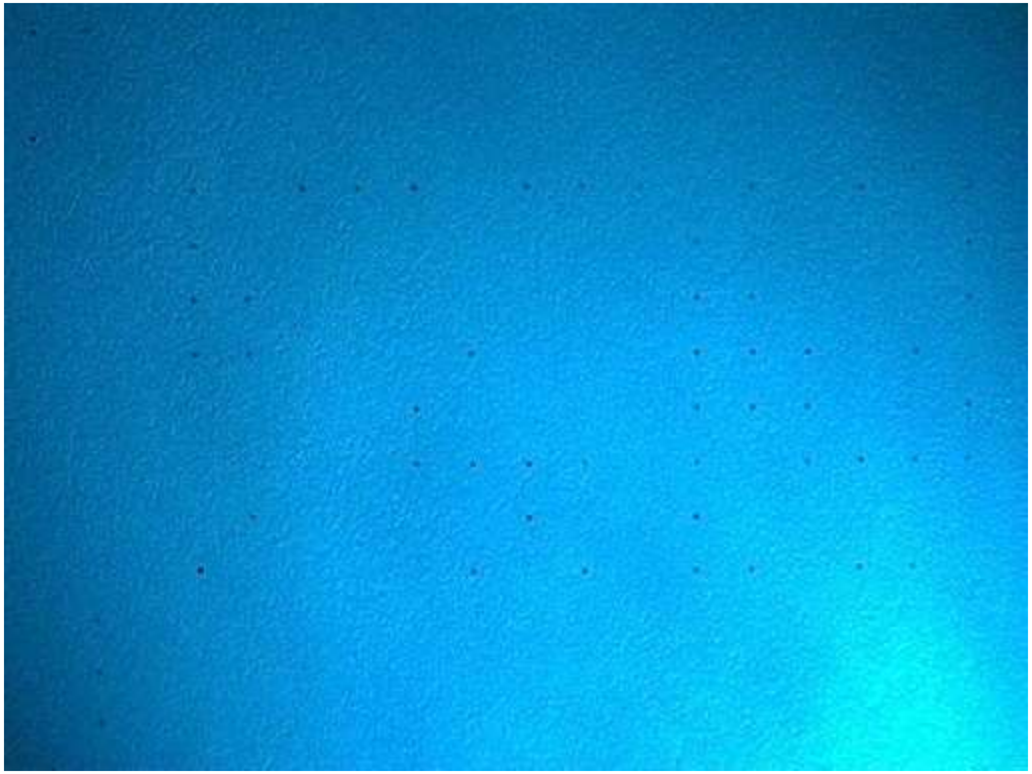
  - Alter the least perceptible bits to avoid detection
    - But these are the same bits targeted by lossy image compression software

Machine ID codes in laser printers

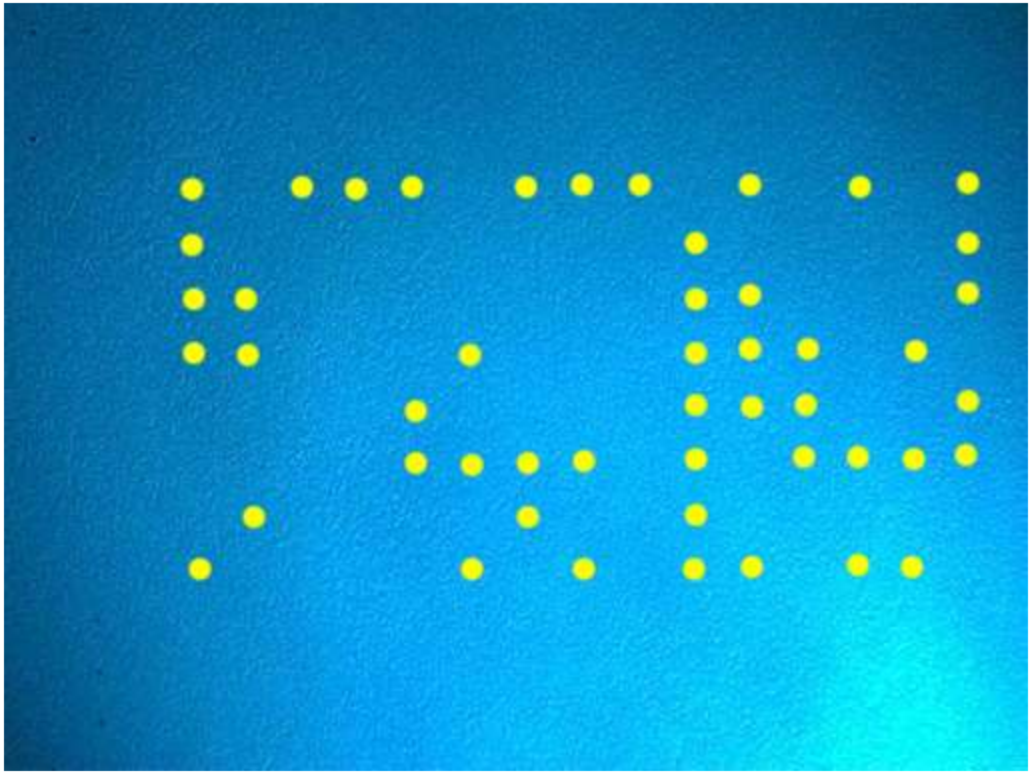


See http://www.eff.org/Privacy/printers/

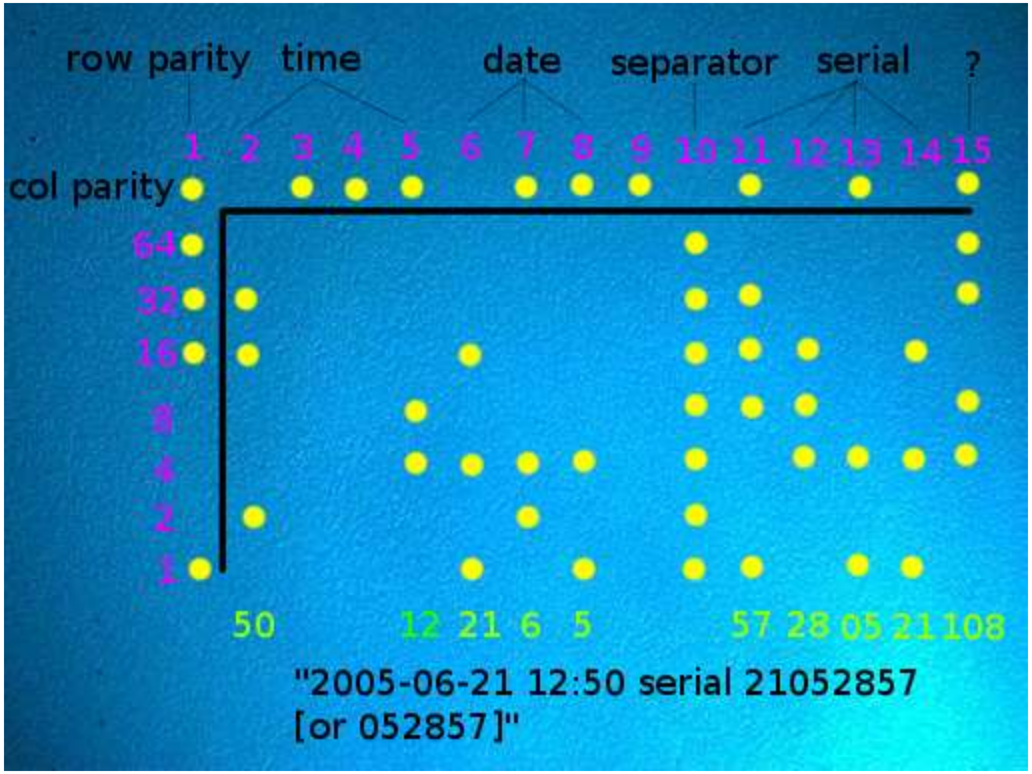
Machine ID codes in laser printers



Machine ID codes in laser printers



Machine ID codes in laser printers



"2005-06-21 12:50 serial 21052857
[or 052857]"

Watermarking vs. Steganography

Both techniques hide a message in data

Goal of steganography

- Intruder cannot detect the message
- Primarily 1:1 communication

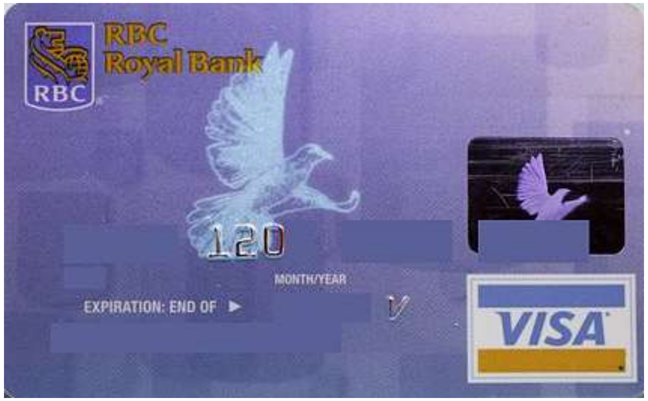
Goal of watermarking

- 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

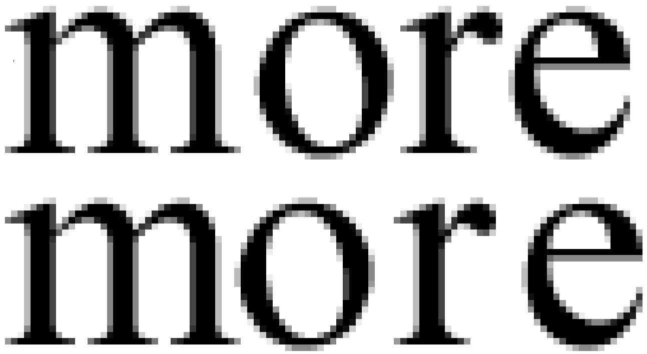
UV Watermarking



Also passports, amusement park re-entry, …

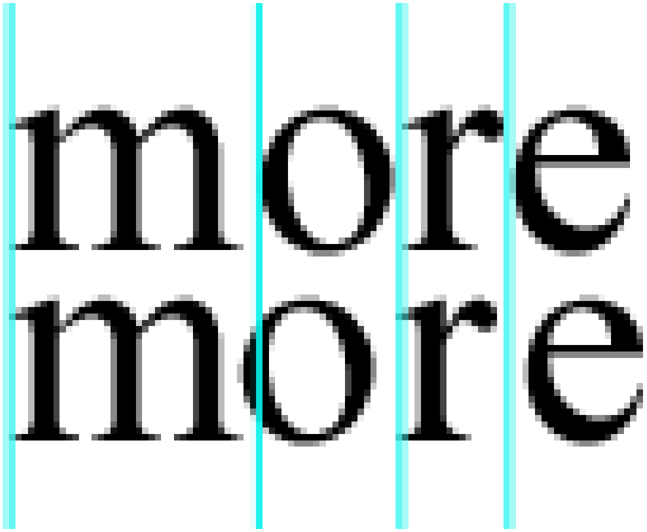
Text

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



Text

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

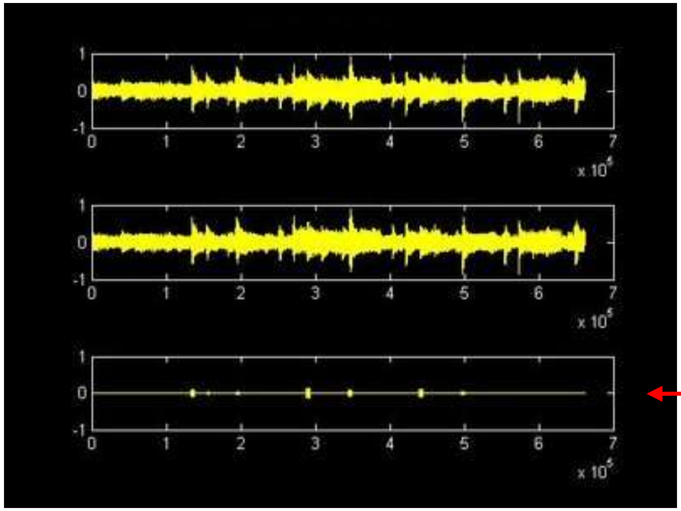


- works only on “images” of text e.g., PDF, postscript

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

Difference

Video

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

The end