# Digital image watermarking in the "real world"

**Adrian Perrig**

**Carnegie Mellon University**

**adrianperrig@innocent.com**

**Joint work with Scott Craver and Fabien Petitcolas**

- Thesis from last lecture: Digital image watermarking can be used for copyright protection of images on the Internet.

- Today's antithesis: Digital image watermarking is not sufficient for copyright protection.

# Outline

- Review of terms and concepts
- Take a close look at different attacks

- The role of watermarking in copyright protection
  - In this lecture, we use watermarking as a mechanism for the automated identification of images and detection of copyright infringement.
  - For copyright protection, it suffices to register the image
  - A judge can decide whether the stolen image corresponds to the registered image
  - Many of the attacks prevent an automated web-spider to detect watermarked images

# Review of watermarking concepts

Watermarks

visible invisible

Verifiability

private public

Watermarking technique

fragile robust

Key

symmetric asymmetric

Original necessary

yes

no

# Problems of digital image copyrighting

- $$$ - big business - big bucks (exp growth)
- Internet spans continents and countries seamlessly
- Digital information is easy to copy
- Hackers are knowledgeable, creative, have lots of time and are numerous
- Many have claimed to have invented an unbreakable algorithm
- Bazillions of attack opportunities
  - Few inventors, many attackers
  - Inventors are few and despair after 3 years
- Human factors:
  - The default user does not understand watermarking
  - Human vision system is very robust to noise in images
  - Used to low quality in images (TV, strong JPEG compression)

# Companies

- Digimarc
- Bluespike
- MediaSec
- Signafy
- Signum (signumtech.com)
- ARIS (musicode.com)
- Intertrust
- But also some of the 2-3 letter companies
  - IBM
  - HP
  - NEC

# Attack classification

- **System architecture failures**
- **Signal diminishment**
- **Image detector failure**
- **Court of law attacks**

- **Note: to illustrate the attacks, Alice is our content creator/owner, Bob is another “good” person and Mallory is the evil attacker.**

# System architecture failures I

- **Protocol attack: Counterfeit original attack**
  - Did you hand in the homework yet?
- **Human factor**
  - Unfamiliarity with watermarking
- **User interface**
  - Having the watermarking tool built into the same program as the image manipulation tools is asking for trouble
  - Should shield the user from error
- **Implementation weaknesses**
  - Digimarc uses public watermarks for authentication of creator and copyright protection
  - Image creator id only has 2-key password
  - Very easy to blackmail another user
  - Debugging tools to change software behavior

# System architecture failures II

- **Web crawler limitations**
  - **Refuse connection to crawlers**
  - **Spoofing, logins, payments**
  - **All the image detector failures we will discuss later**

# Signal diminishment

- Adding noise
- Oracle attack
- “Homemade oracle” attack
- Lossy compression
- Layering attack
- Loosely speaking we want the following property from a watermarking algorithm:
  In order to remove the watermark with sufficiently high probability, the required changes render the image unusable
- But users are usually happy with a low quality level (Jpeg, TV)

# Watermark detector failure

- **Most of these attacks prevent the watermark detector to synchronize with the watermark**
- **Jitter attack**
- **Distortion attack (StirMark) which simulates printing/rescanning**
- **Bandwidth limitation (mosaic attack)**
  - **Watermarking cannot handle small images**
  - **Split images in small pieces (e.g. 100$\times$100)**
- **Java applets/ActiveX controls**
  - **Image displayed with Java applet automatically or after certain actions of the user**
  - **Can even be de-scrambled 'on the fly'**
- **Unanticipated collisions**

# Court of law attacks

- Server in another country “Internet is global but the law isn’t!”
- Problem: web servers do not issue “receipts”
- Spoofing attacks
  - Use impartial witness to prove origin?
  - Blackmail another party to distribute protected content
- Will law enforcement start to download content?

# Open problems

- **Is digital image watermarking the solution for copyright protection? (Since no watermarking algorithm resists all the attacks mentioned)**
- **What about rights management?**
- **Is there a better way to solve the problem of intellectual property protection of digital content?**