Privacy, Identity, Computational Forensics

Prof. Thomas Breuel

other security-related courses

- **89-3552 Security Engineering**
  - protocols, rights, cryptography, security
- **89-4252 Security in Distributed Systems**
  - cryptography, protocols
- **89-4253 Security in Wireless Networks**
  - wireless security, protocols, authetication
- **(anything else?)**

this course

- information theory
- statistical tests
- machine learning
- pattern recognition
- risk estimates
- side channels
- physics and randomness
- psychology and economics

this course

- what is possible?
- how does it work?
- how well does the technology work?
- how can it be used / mis-used?
- what are the challenges?
- what countermeasures are there?

course mechanics

- 2+0 course, once a week
- active participation required
- collect information and contribute
- some computer exercises (Python)

EXAMPLE:

FACE DETECTION & RECOGNITION

face detection & recognition

- **given an image...**
  - find all the faces
  - find the same face in multiple images
  - assign names & identities to those faces
- **security & forensic applications**
  - given an image of a suspect, search mugshot databases
  - spot terrorists in public places
- **but...**

Picasa™ Web Albums | My Photos | People | Explore | Upload

Name tags: **All People**

People: R. **Album:** All Albums

R. (1)
R. Ma ha@cs du

Edit contact
Correct name tag mistakes

Use the checkboxes below to select faces of one person. Select: All None

Enter name: Apply

Suggestions: R. Todd Henry Choose

Don't want to name a face?
Mark selected faces as: Ignore Not a Face

web photo applications

- **consumer application & convenience**
  - people upload their snapshots
  - system identifies faces
  - people label faces of friends with names + email
- **implications?**
  - what can Google use this data for?
  - who else can use this data?
  - how well does it actually work?
  - better than a police mugshot database?

organization

- **course is organized by applications**
- **cross-cutting & recurring**
  - users, targets, measures, counter-measures
  - false alarms and false negatives
  - cost and economic factors
  - statistical analysis & risk analysis
  - pattern recognition & machine learning methods
  - image processing
  - human factors, psychology, perception
  - physics, sensors, randomness

IMAGE-RELATED TOPICS

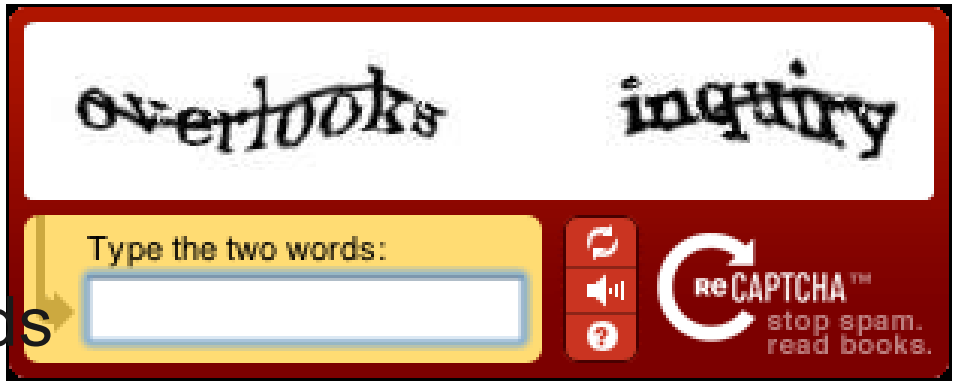
steganography & watermarking

- **hide information in images, text, etc.**

- **applications**
  - covert communications
  - hiding illegal materials
  - tracking / copyright enforcement
  - espionage
  - information leak detection

CAPTCHA, RTTs, Mech. Turks

- **limit access to services to humans**
- **issues**
  - how does it work?
  - who is using it?
  - what is it used for?
  - can it perform useful computation?
  - what are the limitations?
  - how can it be broken?
  - what about accessibility?
  - can we build better ones?
  - visual authentication/passwords



detecting photo manipulation

- **is it real or manipulated?**
- **issues**
  - kinds of manipulations?
  - legal applications?
  - mechanisms of detection?
  - circumvention of detection?
  - cryptographic alternatives?

printer & camera identification

- **find model & identity of device from images**
- **issues**
  - devices: printers, cameras, scanners
  - identification from metadata
  - identification from physical characteristics
  - building forensic device databases
  - identification from social network sites
  - interaction with Flickr, Picasa, etc.
  - countermeasures to identification
  - impersonation & framing

optical document security

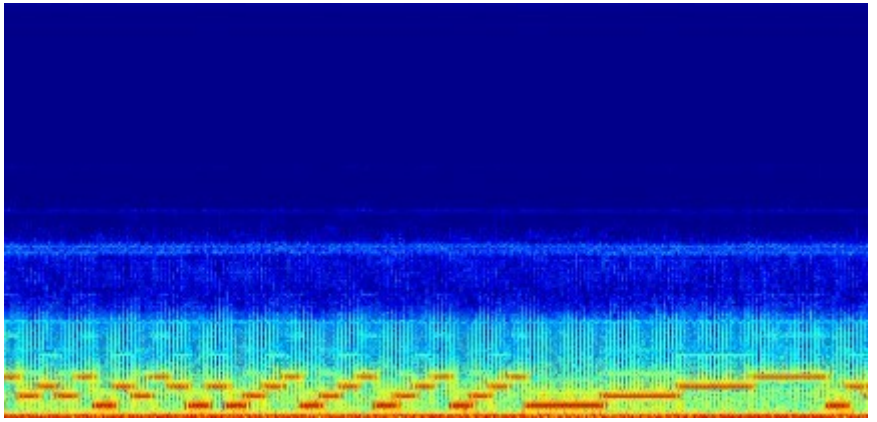
- **traditional means for securing documents**
- **issues**
  - goals and threat models
  - physical randomness
  - special paper, fibers, chads, etc.
  - holograms, gratings, threads, etc.
  - special manufacturing processes
  - visual distinguishability
  - limitations of copiers and printers

biometric identification

- **identification from fingerprints, faces, ...**
- **issues**
  - different types of biometric identifiers
  - physical & biological randomness
  - access control
  - positive and negative errors
  - usability and user acceptance
  - countermeasures
  - sensors
  - cost

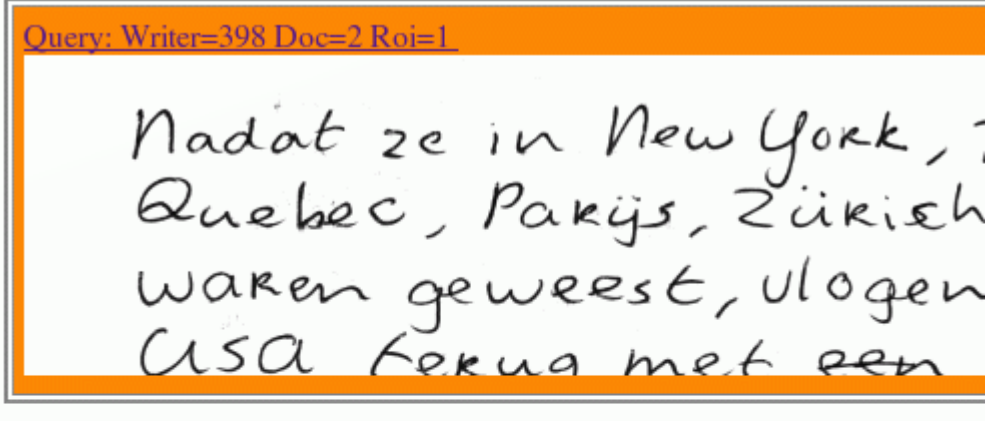
information leaking

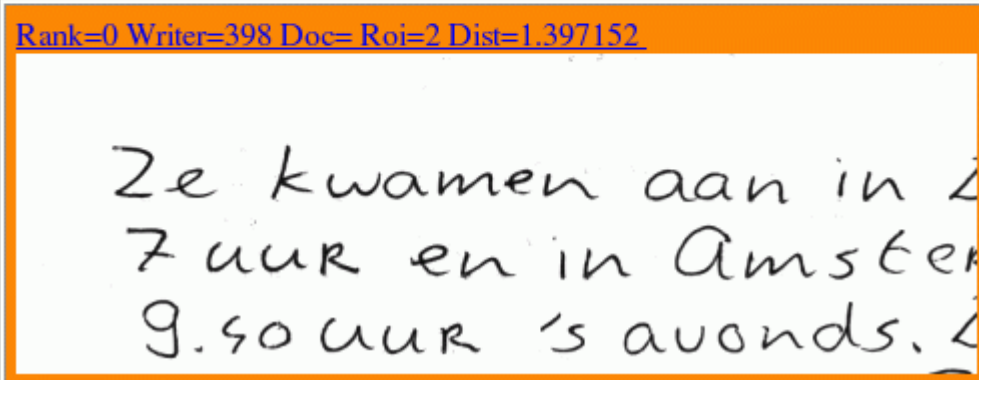
- **“secure” protocols leak information**
- **examples**
  - recovering passwords from encrypted ssh sessions
  - keystrokes from keyboard sounds
  - keystrokes from electrical noise
  - screen content from radio signals
  - distant sound capture with lasers
- **issues**
  - how do these work?
  - how well do they work?
  - countermeasures?



handwriting identification

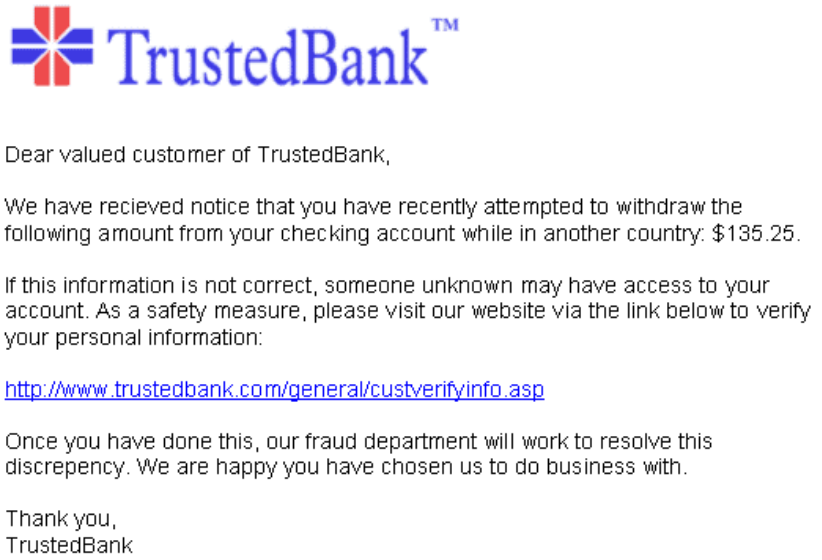
- **sample writing → writer identity**
- **issues**
  - “expert” methods & validation
  - automated approaches
  - forensic applications
  - positive and negative errors
  - validation
  - active countermeasures





phishing

- **tricking people into doing the wrong thing**
- **issues**
  - how are people tricked?
  - psychology and cognitive issues?
  - how does this relate to images?
  - countermeasures?



TrustedBank™

Dear valued customer of TrustedBank,

We have recieved notice that you have recently attempted to withdraw the following amount from your checking account while in another country: $135.25.

If this information is not correct, someone unknown may have access to your account. As a safety measure, please visit our website via the link below to verify your personal information:

http://www.trustedbank.com/general/custverifyinfo.asp

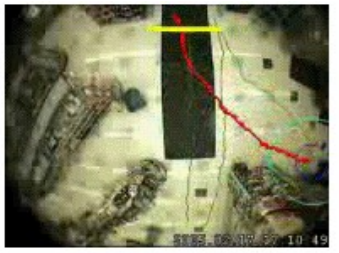
Once you have done this, our fraud department will work to resolve this discrepency. We are happy you have chosen us to do business with.

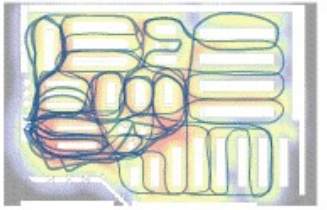
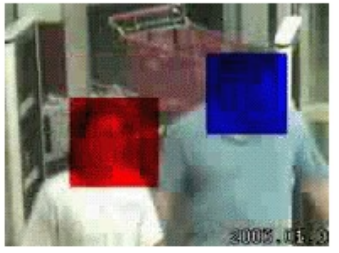
Thank you,
TrustedBank

Member FDIC © 2005 TrustedBank, Inc.

multimedia mining

- **retrieve information from databases/streams**
- **issues**
  - identifying people, locations, activities
  - approaches & their performance
  - pornography detection
  - mining Flickr, YouTube, Facebook
  - real-time surveillance & terrorism de
  - multimedia forensic data analysis
  - database architectures
  - copyright infringement (YouTube)



location & tracking

- **determining location of people & devices**
- **issues**
  - GPS technology
  - localization based on radio signals (WiFi, GSM, ...)
  - localization based on images, sound, etc.
  - building location bases
  - geolocated images
  - forensic image locations
  - device and RFID tracking
  - WiFi-based imaging

security inspection

- **detect weapons & explosives**
- **issues**
  - detecting guns, liquids, explosives, detonators
  - 3D imaging
  - X-ray imaging
  - THz imaging
  - usability issues
  - privacy issues (“screener porn”)

TEXT-RELATED TOPICS

text mining & filtering

- **categorize, detect, identify text content**
- **applications & issues**
  - what is being detected?
  - how is it being detected?
  - named entity detection
  - real-time spam detection & filtering
  - real-time automated censorship
  - CARNIVORE (US, NSA), Great Wall (China), ...
  - censorship circumvention
  - DOS attacks on censorship
  - organizations and political dimension

authorship identification

- **given a text sample, identify its author**
- **issues**
  - historical authorship determination
  - de-anonymization of postings
  - entity-based de-anonymization
  - the AOL query log disaster
  - attribution of forensic evidence, texts
  - impersonation and forgeries
  - identify attributes of the author: skill, bilingualism, etc.

text mining & intelligence analysis

- **support human analysts w/large corpora**
- **applications & issues**
  - philology — TextGrid
  - forensic & intelligence analysis
  - open source analysis
  - scientific text mining
  - text categorization
  - topic identification
  - named entity recognition
  - copy and plagiarism detection

social network & traffic analysis

- **who links/talks to whom?**
- **issues**
  - visualization & support for analysts
  - statistical methods and machine learning
  - ex.: sexual orientation from social networks
  - privacy implications
  - countermeasures
  - balance between privacy and utility

forensic data mining

- **identify individuals/attributes from database**
- **issues**
  - national security, “Rasterfahndung”, insurance applications
  - ad placement and recommender systems (Google)
  - epidemiological & sociological applications
  - statistical and data mining techniques
  - anonymization, reproducibility, validation of results
  - data sources and connecting different data sources
  - errors and their statistical consequences
  - risks, costs, and conflicting interests
  - minimum information necessary to identify a person

NOW?

prepare for next lecture

- try out face recognition in PicasaWeb
- estimate capabilities, error rates
- legitimate uses/services?
- legal issues/rights?
- potential forensic use?
- threat/risk scenarios?