Privacy

Computer Forensics, Budapest 2008

Institute for Information Processing and Microprocessor Technology (FIM)
Johannes Kepler University Linz, Austria

E-Mail: sonntag@fim.uni-linz.ac.at
http://www.fim.uni-linz.ac.at/staff/sonntag.htm
Agenda

- What is privacy?
  - The basic right
  - Giving "consent"
  - Exclusions
- Personal data in computer forensics
  - Web, networks, E-Mail
- Anonymisation proxies
  - Web
  - E-Mail
- Secure data deletion
  - Wiping hard disks
- Countermeasures against data retention
Introduction: Why privacy?

- "Why privacy? I don't have anything to conceal!"
  - In theory, yes, but actually….
    - See film stars: Every photo in a private activity will be published
      - Whether it is scratching your nose or kissing someone
    - Harmless activity can easily be misunderstood or misused!

- A constant thought of "I'm being watched…" builds up
  - Psychologically this produces constant pressure and a general fear
    - "I don't trust you, because you are being watched!"
    - This breeds conformity and prevents any kind of open discussion etc. if taken too far
    - Example: Who will discuss politics if every word is recorded, stored, and later perhaps used against you?
      - Therefore politicians always contradict themselves!

- Constant supervision treats humans as objects
  - Reversal of "innocent until proved"
Why the need for personal data?

- Large, but unfocused, desire for privacy by individuals
  → “The right to be left alone”
- Large desire for information by companies
  → Know your customers, advertising, credit rating, …

Some balance must be found!

- Otherwise there will be:
  → The danger of secretly gathering data
  → The danger of exchanging and correlating data uncontrolled
  → No advantages of personalization
  → No advantages of not requiring standard information
  → …
Data Protection: Protection against disclosure
   → Data should be kept secret

(Data) Privacy = Data Protection

Data Security: Protection against loss
   → Data should be available (to the subject and the owner)

Data Safety = Data Security

Both aspects are important
   → Here only the first one is discussed!
Privacy vs. terrorism

- But in some cases monitoring is necessary
  - This has already been acknowledged by privacy laws
  - The important discussion is: Where to draw the line!
- Terrorism is a very "public" crime: Although the number of people dying by it in western countries is negligible compared to car accidents, it is an excellent "reason"
  - Nobody fears being hit by a car,
  - but (almost) everyone is in panic of bombs!
- Terrorism is a problem, as "modern" terrorism is almost impossible to stop by surveillance. It only helps afterwards to identify terrorists and perhaps some of their associates
  - This is still important, but one step less than prevention
  - This area is currently hotly disputed, and politics (not necessarily the police!) request lots of additional options
The basic right to privacy

- The right to privacy is "the right to be let alone"
  - Not everything a person does may be observed, noted, used, stored, calculated with etc.
- This includes the prohibition to obtain personal data!
  - The problem is not necessarily what happens with the data, but that data exists at all: It might be (un-)intentionally disclosed; and if data exists it will be used sometimes!
    - See the highway toll in Germany as an example
- Sometimes personal data is "known" inevitably
  - Example: Doctor's secretaries/aides
  - Then privacy refers to the prohibition to disclose the information to third persons or use it for any other task

Please note: From here on the content refers to the EU privacy directive!
Who is protected?

- EU directive: **Only** natural persons
  - Austria: Extended to legal persons
  - The intention is to protect humans from everything/-one else
    - Children in relation to their parents
    - Employees in relation to their employing company
  - Legal entities are often protected only to a lesser degree
    - See e.g. publishing financial data; or environmental pollution
    - **Included** in directive on privacy and el. communications!

- Only persons which are identified or identifiable
  - If nobody can say who the person is the data is related to, there is no danger at all (purely statistical data)
  - Identification can be possible directly or indirectly
    - E.g. one/more factors specific to physical, physiological, mental, economic, cultural, social identity
    - "The blonde girl working in the accounting department", if there is only one a) young woman, with b) blond hair, c) in that department
What is protected?

- All data relating to a protected person
  - Example: Hair colour, voice, letters, personal habits or preferences, income, sexual orientation, last breakfast meal, creditworthiness, ...

- Special protection exists for more "dangerous" data:
  - "Sensitive" data: Closed list
    - Racial/ethnic origin, political opinion, religious/philosophical beliefs, trade-union membership, health, sex life
  - "Criminal" data: Closed list
    - Offences, criminal convictions, security measures
      - Does NOT refer to administrative sanctions or judgements in civil cases (national law may include them, however!)
  - These two areas are more strongly restricted, but numerous exceptions are still possible (see later)
    - Normal data: "public interest"
    - Sensitive data: "substantial public interest"
What is protected?

- Only data that is processed
  - Gathered, related to other, transferred, ...
  - But **NOT** the data as such!

- “Public” data **might** still be protected!
  - Especially if known only to a restricted public

- Data **must** be either
  - automatically processed, or
    - »Computer systems in any form
  - contained (or intended to be contained) in a filing system
    - »Criteria related to individuals necessary
    - »Unimportant: local or distributed / functionally or geographically
    - »E. g. Database, filing cabinet with index

- **NOT** included are unordered collections
  - When there is at least one criteria for searching the content easily (not just serial exhaustive search), it is protected!
"Consent" in the context of privacy

- Must be "informed consent"
- This includes three aspects:
  - Must be given freely ("Freedom")
  - Must be specific ("Specificity")
  - Subject must be informed ("Information")
- Can be given explicitly or implied
  - What is needed depends on the category of data
    » "Normal" data: Implied consent sufficient
    » "Sensitive" data: Consent must be given explicitly
  - Usually there is no need for consent in "writing"
    » At least not in the EU directive
    » Germany previously required "written consent", which was a problem in the Internet, although el. signatures did exist then
      – But nobody used them, so this was abolished!
- No duress or compulsion
  - Denial of contract (if not unethical) possible if not given!
    » BUT: Effective monopolies; e. g. banks???
    » In practice quite a lot is possible under such a condition
- But these are necessary conditions for every legal act?!?
  - There must be a bit more freedom!
- One typical example are work contracts
  - In the contract usually quite a lot of conditions can be added
    » Everyone is free to accept this contract or decline
      – This is perhaps not that true in practice….
  - But for an employee there is almost no possibility to give valid consent to his employer later!
    » "You **will** allow this or I'll **sack** you!"…
Information

- Information necessary on
  - That some personal data is used
    » "We will collect, store, …. your personal data …"
  - What data is used
    » "We collect your IP address, web sites visited, and all information on the forms filled in"
  - Who is the person using it
    » "We are the ACME Inc."
  - Whom it will be transferred to (if applicable)
    » NOT “to all members of our company group”

- Especially important for implied consent
  - Consent is only possible to the things actually disclosed!
• Consent cannot be given for multiple applications
  → Only (a list of) single applications, but not a "general" consent
• Specificity means:
  → For a certain purpose: A closed list/described set
    » NOT “we are allowed to do with it what we want”
    » This is the most important part!
      – Example: "Advertisements" is not specific enough
    » However, no absolutely closed list is required
      – "Marketing our own products" could be sufficient
  → For a certain controller
    » NOT “we may transfer it to everyone we like”
  → Of certain data
    » NOT "whatever we know or find out about you"
Exclusions: Overview

- The basic right prohibits any use of personal data
  - See above: This will not work in society
- Several exclusions exist, when personal data may be collected, used, stored etc.
  - Typically, transferring the data is much more restricted!
  - Fewer exclusions exist for the more "dangerous" subsets of data: sensitive and criminal data
- In the EU directive the exclusions are very general
  - National law can either define them in more detail, like in Austria, or leave it up to the courts
- In general, there is a weighing of interests between the person the data is about the person wanting to use it
  - Some decisions of this weighing has been included in the directive as a pre-determined result!
Normal data may only be processed if

(1)

- the subject has unambiguously given consent
  - See previous definition of consent!
    - "Unambiguous" → Implied consent is possible here
  - Everyone can do with his data as he wants
    - The freedom not to use the protection of the law
    - See e.g. television talkshows!

- it is necessary for the performance of a contract
  - Data subject must be party to contract, or
  - for taking steps at request of subject prior to contract
    - E.g. checking creditworthiness, calculating shipping costs, …
  - Otherwise this could be used as a right of withdrawal!
    - If later you do not want the contract any more and prohibit the seller to use your address ⇒ He couldn’t ship the goods!
Normal data may only be processed if

1. it is necessary for compliance with a legal obligation
   - Obligation of the controller, i.e. the one processing the data!
   - Examples: archiving of invoices, processing data of the employee by the employer (holidays, payment, ...)

2. it is necessary to protect vital interests of the subject
   - E.g. looking up her own blood group on serious injuries
   - “Vital” must be seen narrowly
     » “Of interest” or “possibly beneficial” is not enough!

3. it is necessary for tasks of public interest/official authority
   - To avoid having to explicitly grant all processing by law
   - Must be an important or indispensable requirement, not just a reduction of work

4. it is necessary for legitimate interests of the controller, third parties, or those to whom data is disclosed
   - EXCEPT where the interests of subject are stronger!
Weighing of interests required

- This is an "opening clause": You may do whatever you want with any "normal" personal data, but you need to have:
  - Some interests: Easy!
  - They must be legitimate: Usually no problem!
  - They must outweigh the interest of the person to keep the data private: Most important and typically difficult aspect!

- Examples:
  - Vital interests of thirds: Searching DBs to find suitable blood donors
    - To contact them to ask, whether they would be willing to donate blood
  - Required for pursuing a claim before public authorities
  - Cooperation through official channels to improve public admin.

- May **not** be just a monetary comparison
  - Gain for processor vs. damage to subject → Always insufficient!

- General clause for all other uses!
Sensitive data may only be processed if

1. the data subject has given explicit consent
   - Countries can define some areas, where even consent is not enough, i.e. where the person is protect from itself!

2. processing is necessary for carrying out obligations/specific rights of the controller in employment law
   - If this is authorized by national law
   - Adequate safeguards must be ensured
   - Example: Accounting includes health information
     » AT: Trade-union membership fee is partly collected by employer!

3. processing is necessary to protect the vital interests of the subject or another person
   - Only if subject is physically/legally incapable of giving consent
     » Otherwise: The subject must be asked!
   - No denying possible regarding vital interests of others!
Sensitive data may only be processed if

- processing occurs by a foundation, association, ... with a political, philosophical, religious or trade-union aim for their members or persons with regular contact connected with their purpose
  - I.e., churches may have lists of members and supporters
  - Only for legitimate activities and with appropriate guarantees
  - This data may not be disclosed to thirds without consent!
- the processing of data manifestly made public by the subject
  - After a "coming-out" the sexual orientation may be stored
- the processing is necessary for the establishment, exercise or defense of legal claims
  - You may use personal data in courts to prove your case
Sensitive data may only be processed if (3)

- the processing regards preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services
  - Data must be processed by a health professional with an obligation of professional secrecy (or persons with equivalent obligation of secrecy)
- Other national legal exemptions with suitable safeguards are possible for reasons of substantial public interest
  - Examples: Private use, scientific research, statistics, informing the data subject, catastrophies etc.
Almost all data in computer forensics is personal data

- This is typically the interesting part: Data as evidence what a certain person did do (or did not do!)

So care must be taken to only search for/extract/recreate data for which there is sufficient legal reason

- Otherwise sanctions may be imposed
  » Including criminal proceedings!
- Attention: Several tools used for forensics are "dangerous"
  » Already the simple possession may be illegal if combined with a certain intention (even more its distribution, making available, …)

Obtaining permission is therefore paramount

- Either from all persons from which data may be on
  » Attention: E.g. E-Mail → Consent of recipient and sender!
- Or from someone else, for instance the court
Files may contain any, including sensitive, personal data

- So potentially a hard-disk as a complete unit is subject to the strongest restrictions
- Inspecting a file therefore needs also the strongest exception
  » However, the file name may be a guide for the content

Attributes can also contain personal data:

- Who created/accessed the file (last)
- When was the file created/accessed

Restrictions are possible to certain shares, partitions etc.

- If the owner of this partition gives consent → no problem
- This does not apply to partition slack or general partitions!
  » Boot partition, swap partitions, …

Not all data is personal data: Program code, OS

- But: Configuration files (Registry) etc. do contain such!
Personal data in network transmissions

- Observing network data also refers to personal data
  - Typically the content of the communication
    » Files transferred, E-Mails being sent/received, ….
  - The recipient/sender address
    » IP addresses can be personal data
      – WLAN: Typically only local, so with other content, DHCP server etc.
        the person can be identified
      → Almost everything becomes personal data!

- But there is also technical data
  - Protocol overhead, system communication, etc.

- Criminal sanctions of intercepting communication exist, too!
  - Convention on Cybercrime, national laws, …
E-Mails are very typical personal data
  → Both recipient and sender need to be protected

Personal data:
  → The actual textual content (or images, …)
  → The subject line
  → The recipient/sender address
  → The sender IP-address
    » Provides information on the location of the sending
      – Not necessarily where the E-Mail was written!
  → The time stamp(s): When the E-Mail was sent
  → Other headers: The software used, …

E-Mail, subject, and addresses can even be sensitive data
  → Example: helpline@drugabuse.com, "The pains in my leg", …
Basic principle of anonymisation is routing the traffic across one or several different computers, so it appears to be coming from there instead of the real origin
→ I.e., hiding your IP address!
→ Additionally, there no logs on the "real" source may be kept

Problem: Communication must be secured, otherwise interception on the source side provides all the information!
→ Solution: Encrypted communication with the proxy and secure identification of the proxy

Problem: Correlating input and output still possible
→ Solutions: Random delay, networks of proxies
   » Requires lots of users to prevent this ("hiding in the masses")!

Problem: The fact that a proxy is used can be interesting
→ Solution: Currently none (at least useful; → steganography)!
Web surfing anonymisation

- Problems:
  - Delays are not possible – "Realtime" forwarding necessary
  - Format of HTML requests is very simple and well-known
    » Starting text is known, size of request can provide information
  - High throughput needed (binary downloads!)
- Security: The anonymisation does not apply to the proxy!
  - It can log all usernames, passwords, create copies of files, …
    » Cascading: Only the first and last one; others may be encrypted
- Locking out: Some servers reject requests from known anonymisation proxies!
  - To avoid legal problems (or spamming!)
TOR (The Onion Router)

- TOR is a free TCP proxy
  - All TCP traffic can be anonymized, not only web browsing!
    » But E-Mail is usually forbidden, to prevent SPAM!
- How does it work:
  - Each connection takes a random way over several nodes
    » The next connection may use a different route!
  - Each hop is encrypted separately
- Problems:
  - Intermediate proxies are unknown: Whether they are trustworthy (no logging) or not, is unknown
  - DNS is not TCP, but UDP → No anonymisation
    » DNS for "google.hu" → later anonymous request is known!
      – Use the tool "Privoxy" together with TOR; or the 0.2 branch
  - Traffic analysis: A paper showed, that even with only a partial view of the network, anonymisation can be reduced/broken
TOR (The Onion Router)

How Tor Works: 1

Step 1: Alice's Tor client obtains a list of Tor nodes from a directory server.

How Tor Works: 2

Step 2: Alice's Tor client picks a random path to the destination server. Green links are encrypted, red links are in the clear.
JonDonym/JAP/AN.ON

- Commercial successor of the Java anonymisation proxy JAP
- Consists of several features:
  - Mixing: Several proxies after each other, randomly selected
    - Also mixes/combines the requests of several users
  - Mix cascades: Proxies from different operators are used
    - Only a single one must be trusted to be anonymous
    - The proxies are known to the end user, who can also select them
    - In different countries, so court orders to log traffic of certain users will not work
      - Occurred with the predecessor JAP in Germany!
- Client program needed: Redirects the requests to the proxies and encrypts them
- Special functionality to avoid blocking the service:
  - Other "normal" users may act as forwarders to the network
E-Mail anonymisation

● Intention is especially hiding the sender address, not only the IP it was sent from
  → Chaining remailers increases privacy
  → Encryption can be used to render eavesdropping useless
    » Encryption can be hop-by-hop or layered
  → Random delays are possible here (asynchronous comm. !)

● Problems:
  → Length attacks (correlating input and output length) possible
    » 756 Bytes in and 756 Bytes out → Same message
      – Random padding can be used
  → Depending on the system, no answers are possible
    » Some systems contain reference lists (Sender ⇔ pseudonym)
      – These are then in danger of break-ins or official searches
  → E-Mail content can render the anonymisation meaningless
    » "Send products to …", signatures, metadata in attached files etc.
Secure deletion of data

- Possible according to various intentions:
  - Just not visible: Delete with any file program
  - Actually removed: Overwrite content with special programs
  - Removed without traces: Overwrite also directory and slack
    » Better: Also overwrite remapped sectors
  - Really deleted: Remove all traces of the previous magnetic orientation on the disk
    » Degaussing (difficult for modern disks), physical destruction

- RAM content can also be recovered
  - The longer a memory cell holds the same value, the better and the longer it will retain it after power-off

- CD-Roms, tapes: Shredding is the best method

- Note: Usually it is detectable, **that** a drive was wiped!
Wiping disks

- To avoid "normal" recovery by software tools, overwriting all data on the disk a single time is sufficient
  - Magnetic Force Microscopy (MFM), etc. → Much more difficult to protect against, but also rather rare and expensive

- Different approaches to wiping exist:
  » Attention: "All bits" need not be the same on physical surface!
  » Run-length-limited encoding (or others) is sometimes used!
  - Single pass: Random data, all zeros, or all ones
  - Triple pass: All Zeros, all one, random data
    » DoD standard 5220.22 M ("NISPOM")
  - Seven passes: 1, 0, 1, 0, 1, 0, random
    » Canadian standard
  - 35 passes: 4 random, 27 special for RLL, 4 random
    » "Gutmann standard"
Selecting the correct privacy level

- Privacy can be enhanced significantly in various ways
  → But they are typically costly (money, time, effort, …)
- So not everything possible makes sense
- Typical tradeoffs include:
  → Use secure wiping of disks with several passes
    » Everything more is probably not useful: Are your systems so secure that there is no danger of infiltration by the secret service through other avenues (trojans, bribes, etc.)?
    » Important for private persons and companies!
  → There is no need for E-Mail anonymisation
    » Only special cases: Tipping of the press, repressive countries,…
  → Web anonymisation might be useful in rare cases
    » Difficult is not to forget it: A single time without → No anonymity!
    » In general, there should be no need!
Data retention

- Data retention according to the EU directive is rather "weak"
- It ensures the identifiability if the IP address is known
  - Through the provider the computer can be identified
    - Or at least the calling number for dial-in
      - Which must be identifiable too!
  - Not necessarily the actual user, i.e. within companies (NAT!)
- Internet E-Mail and Telephony
  - Information to retain:
    - Sender and recipient (caller and callee) are identified
    - Date and time of checking/sending a mail respectively logging into the VoIP system are stored
    - The Internet service used (i.e. provider, kind of service)
  - Both is possible through the E-Mail/VoIP provider
    - But only this provider must store, not the access provider!
Countermeasures against data retention

- Several general approaches exist:
  - Hide the IP address
    » Impossible: Every computer MUST have one!
    » But we can make it look like coming from a different one …
  - Use "anonymous" sender/recipient IDs for E-Mail and VoIP
    » Sender is no problem: Leave it out or invent it!
    » Recipient: Not really possible; but we might masquerade …
  - Use providers data retention doesn't apply to
    » The EU directive applies to the EU only …
  - "Hide" the communication from the retention
    » E-Mail and VoIP are the only ones under surveillance
    » So use different ones!

- On the following pages various concrete examples are given
  - Other are possible!
  - These are just a few trivial ones!
Non-standard ports

- SMTP and VoIP traffic uses standardized ports
  - But they can be changed manually to any other number!
- Problem: This only works within a closed user group
  - No communication to or from "outsiders"
- Problem: These protocols can easily be recognized according to their content (HELLO - handshakes)
  - But this would mean inspecting the content!
    » Typically illegal
    » Compared to just logging the "normal" ports this requires an extreme increase in computing power!
      - Every single TCP connection must be checked!
- Note: This helps against "monitoring" E-Mail/VoIP by the access provider, which is NOT required!
  - The closed group \textbf{MIGHT (legally!)} have to retain the data …
Alternative software can be used:

- This might still qualify as "E-Mail" or "Internet telephony", but with direct communication between the participants there is no provider who would have to retain this information …
  - Might also be excluded, as only defined protocols are probably stated to be monitored in the national laws

Note: Chat is not E-Mail and not Internet telephony!

- No obligation for data retention at all …

Problem:

- Not trivial to create
  - But only some programming skills are required
- Complete traffic analysis would be necessary to detect
Encryption

- Use encryption to communicate with other persons
  - This only works if there is no intermediate provider
  - Direct communication to the recipient or outside the EU
  - Result: No identification of the content possible at all
    » Only that a certain communication took place ➔ Alternative ports!

- Problems:
  - Online searches can subvert this, as they are before/after the en-/decryption takes place
Conclusions

- Privacy is an important aspect in a free society
- Computer forensics must take great care, as very often the intention is to uncover personal data, the person it relates to explicitly wanted to keep secret
  - Verification of the "permission" is very important
- Data retention will come to a certain degree
  - But it is unrealistic that it will ever reach its goal: Terrorism!
  - However, even very small misdemeanours could be included
  - Additionally, data collected is data misused
- So there is sufficient reason for everyone to take some care and perhaps try to reduce the personal "footprint"!
Questions?

Thank you for your attention!