



Current legal developments

Security and Privacy
VSE Prag, 9 - 13.6.2008

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>



- Personal data and computer forensics
 - Disks, networks, ...
- Anonymisation
 - TOR
 - JonDonym
- Secure data deletion
- Countering data retention



Personal data in computer forensics

- 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



Personal data on hard disk

- 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, ...



Personal data in E-Mails

- E-Mails are a very typical example of 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
 - » and perhaps the location: Timezone!
 - Other headers: Software used, ...
- E-Mail, subject, and addresses can even be sensitive data
 - Example: `helpline@drugabuse.com`, "The pains in my leg", ...



Anonymisation proxies

- 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")!
- Note: The fact that a proxy **is** used can be interesting **itself**
 - 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!)

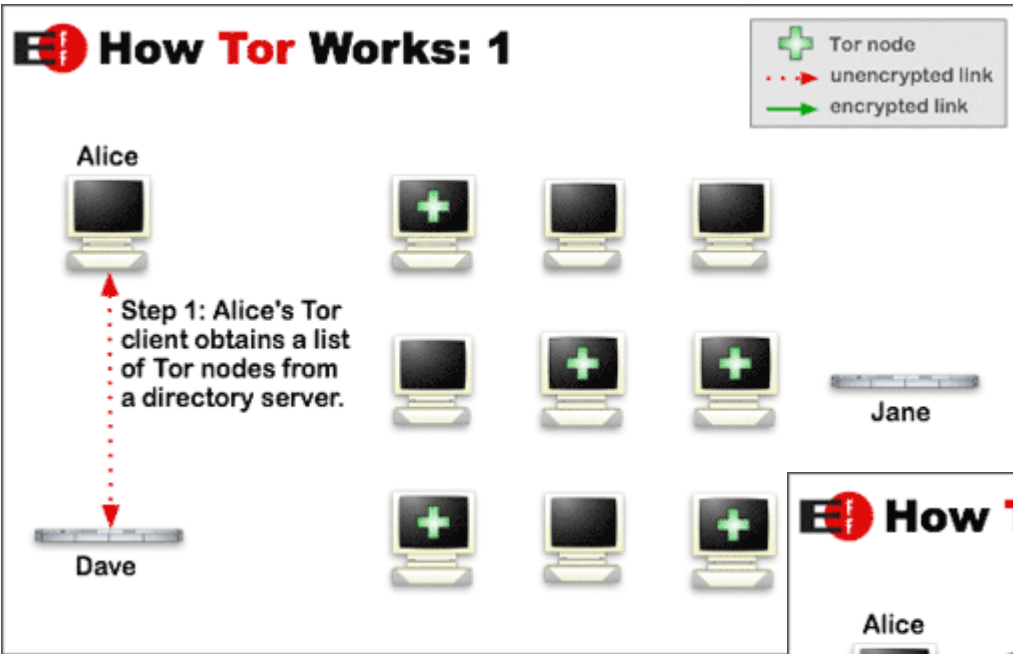


TOR (The Onion Router)

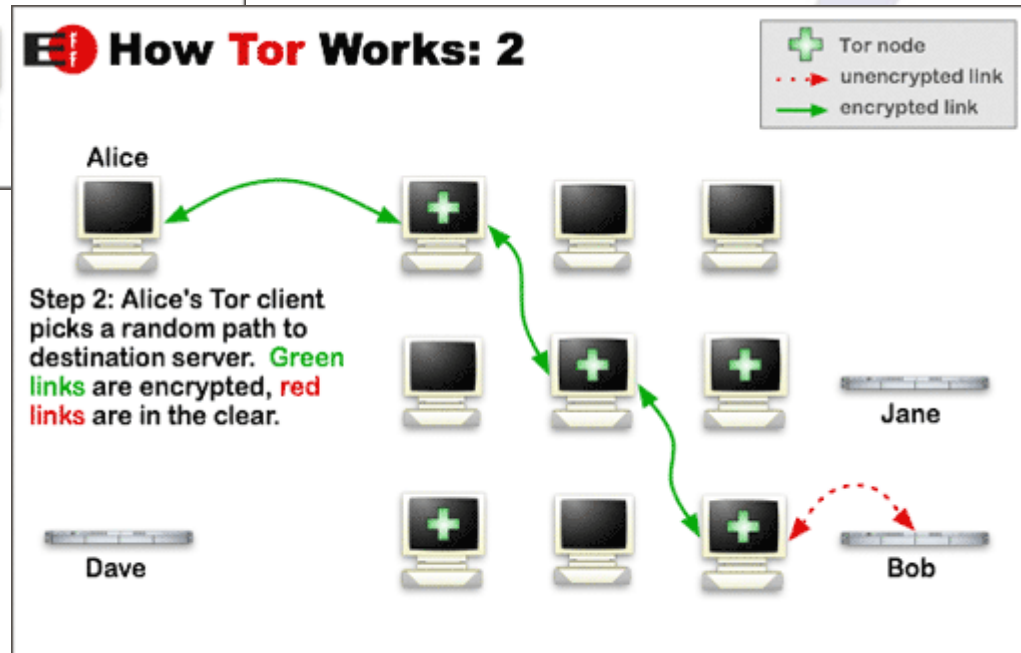
- TOR is a free TCP proxy
 - All TCP traffic can be anonymized, not only web browsing!
 - » But E-Mail is usually forbidden to prevent SPAM!
- How does it work:
 - Each connection takes a random way over several nodes
 - » The next connection may use a different route!
 - Each hop is encrypted separately
- 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

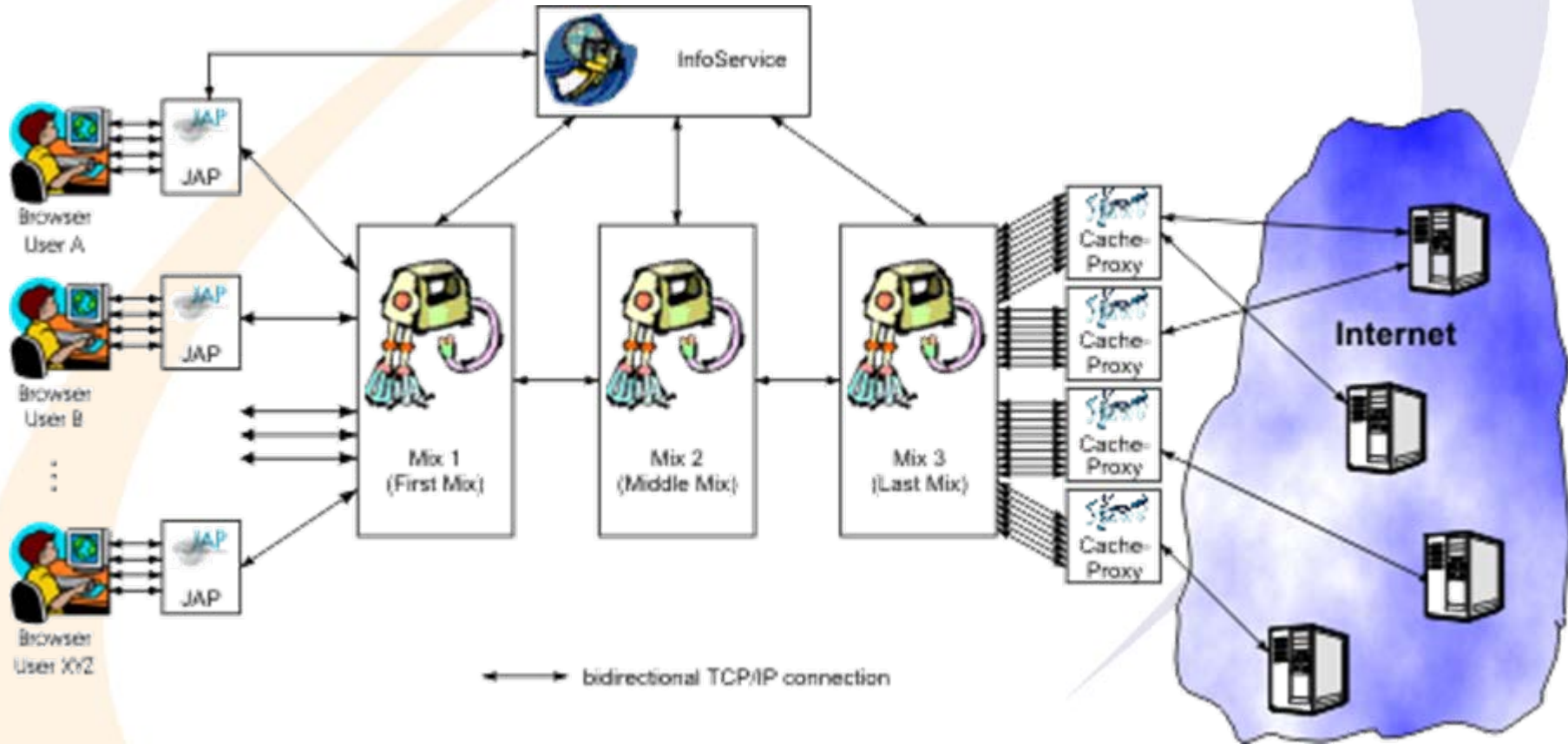


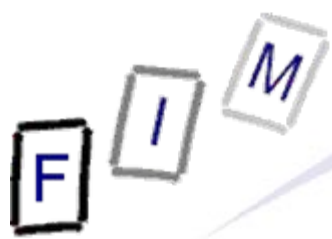
How Tor Works: 2





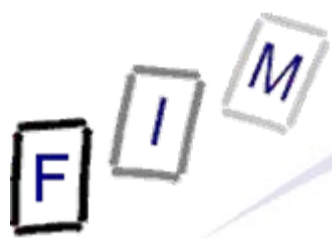
- Commercial successor of the Java anonymisation proxy JAP
- Consists of several features:
 - **Mixing:** Several proxies after each other, randomly selected
 - » Also mixes/combines the requests of several users
 - **Mix cascades:** Proxies from different operators are used
 - » Only a single one must 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
 - » Even longer when (strongly) cooled
- CD-ROMs, DVDs, tapes: Shredding is the best method
- Note: Usually it is detectable, **that** a drive was wiped!



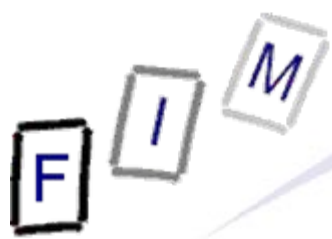
Wiping disks

- To avoid "normal" recovery by software tools, overwriting all data on the disk a single time is sufficient
 - Magnetic Force Microscopy (MFM), etc. may be able to recover traces
 - » Source: No exact positioning (small areas to the side), deeper down below the surface, small differences in magnetism, ...
 - 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 the logical and the physical layer!
 - See especially USB sticks (i.e. everything with Flash memory): Remapping sectors to reduce the wear of the cells
 - » Run-length-limited encoding (or others) is sometimes used!
 - Today not any more



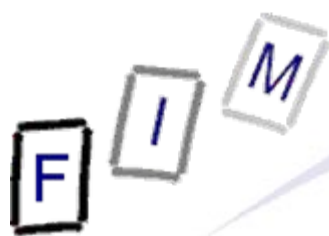
Wiping passes

- Wiping software usually supports one/more of these:
 - Single pass: Random data, all zeros, or all ones
 - » Usually sufficient for companies
 - » No software-only solution can recreate the data
 - 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
 - » Improved security against MFM
 - 35 passes: 4 random, 27 special for RLL, 4 random
 - » "Gutmann standard"
 - » Especially for RLL hard disks (today rarely used)
 - Not very useful today anymore
 - » Very time-consuming!



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 one/three 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!
 - Before selling/throwing away old media
 - There is no need for E-Mail anonymisation
 - » Only special cases: Tipping off the press, repressive countries, ...
 - Web anonymisation might be useful in rare cases
 - » Difficulty is not to forget it: A single time without → No anonymity!
 - » In general, there should be no need!



- 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, like custom protocols!
- 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 would have to be checked!
- Note: This helps against "monitoring" E-Mail/VoIP by the access provider, which is NOT required!
 - The closed group **MIGHT (legally!)** have to retain the data ...

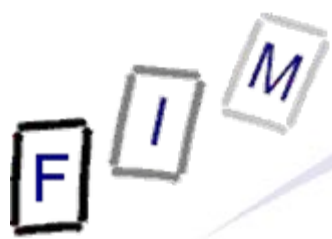


Alternative software

- 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



- 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
 - » They will even allow observing the private key



- 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 = data misused at some time
- So there is sufficient reason for everyone to take some care and perhaps try to reduce the personal "footprint"!

F I M

Questions?

Thank you for your attention!