



Mag. iur. Dr. techn. Michael Sonntag

Ensuring privacy

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>

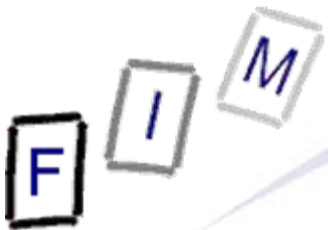


- Personal data and computer forensics
 - Disks, networks, ...
- Anonymisation
 - TOR
 - JonDonym
- Secure data deletion
- Countering data retention



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, which data may be about
 - » Attention: E-Mail → Perhaps consent of recipient **and sender!**
 - Or from someone else, for instance the court
 - Or some other justification (→ weighing of interests!)



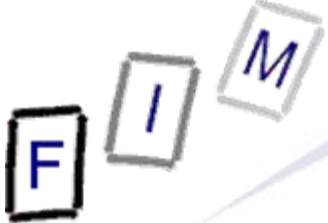
Personal data on hard disk

- 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 (+ login times → user)
- 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, ... → System owner
- 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: Usually only local, so with other data (DHCP server etc.) the person can be identified
 - Almost everything becomes personal data!
 - Germany: Problem for webserver logs
 - » No storing of IP addresses because of privacy (disputed!)
- 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 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 timestamp(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", ...



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, no logs on the "real" source may be kept there
- Problem: Communication must be secured, otherwise interception on the source side provides all the information!
 - Solution: Encrypted communication with the proxy and its secure identification
- Problem: Correlating input and output still possible
 - Solutions: Random delays, network of proxies
 - » Requires lots of users to prevent this ("hiding in the masses")!
 - If large parts of the network are under control, anonymisation can be broken



Anonymisation proxies

- Problem: The fact **that** a proxy is used can be interesting
 - This is possible when listening in on the sender
 - » IP addresses of proxies are known
 - This is possible for the recipient
 - » Some web servers refuse to serve anonymous proxies
 - » Many mail server refuse pass on mails from proxies (→ SPAM!)
 - » Legal reasons may also exist
 - Solution: Currently none (at least useful; → steganography)!
- Problem: You have to trust the proxy
 - If he says he doesn't keep logs, then this is not necessarily the case
 - Based on court orders such proxies **did** keep logs on certain persons – but this requires prior knowledge and suspicion!



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
 - E.g. file upload, comparing to known URLs
 - High throughput needed (binary downloads!)
- Security: The anonymisation does **not** apply to the **proxy!**
 - It can log all usernames, passwords, create copies of files, ...
 - » Note: Data retention in Germany requires this!
 - Cascading as a countermeasure
 - » First one: Knows the source
 - Might be possible to claim "I'm just another one in the proxy chain"
 - » Last one: Must deliver content in cleartext → TLS is important!
 - » All others **may** be encrypted (and typically are!)



TOR (The Onion Router)

- TOR is a free TCP proxy
 - All TCP traffic can be anonymized, not only web browsing!
 - » But E-Mail usually forbidden (proxy can decide what to accept)
- 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
- See also the tool "Tork":
 - Based on TOR (UI/configuration helper for it)
 - Allows in-/excluding servers/countries from the proxy network
 - Supports web-browsing, E-Mail, IRC



TOR (The Onion Router)

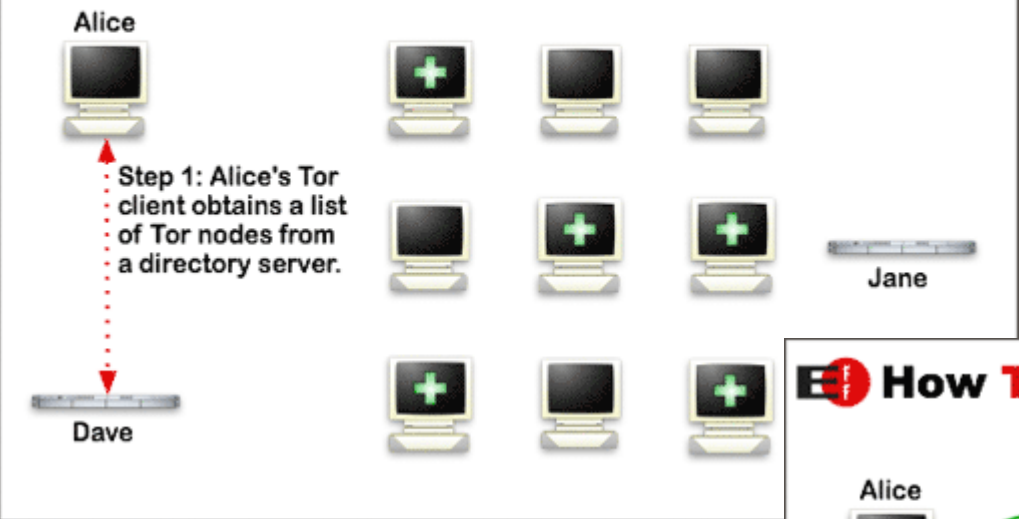
- Problems:

- The last hop has (and always must have) access to cleartext!
 - » Unless using TLS or something similar
 - Log-ins and passwords of ≈100 embassies sniffed by adding 5 exit nodes (which anyone can add!)
 - TLS proxies do exist (man-in-the-middle attacks), as certificate warnings are usually ignored by users
 - » Some nodes only forward the unencrypted protocols ...
 - Government agencies **might** be involved!
 - » Any proxy can modify the data which it forwards ...
- DNS is not TCP but UDP → No anonymisation
 - » DNS for "google.at" → Subsequent anon. request is known!
 - Use additionally the tool "Privoxy"; or the (current) 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)

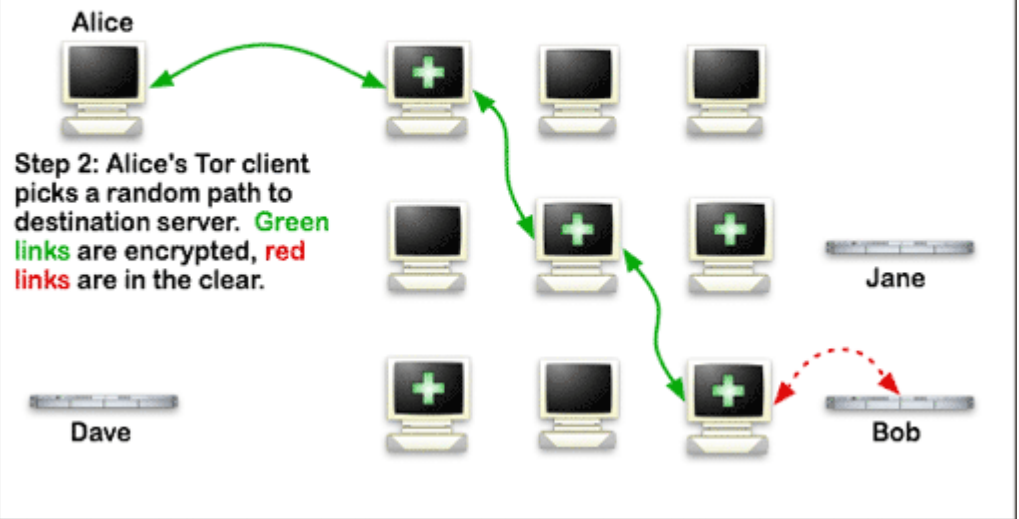
How Tor Works: 1

-  Tor node
-  unencrypted link
-  encrypted link



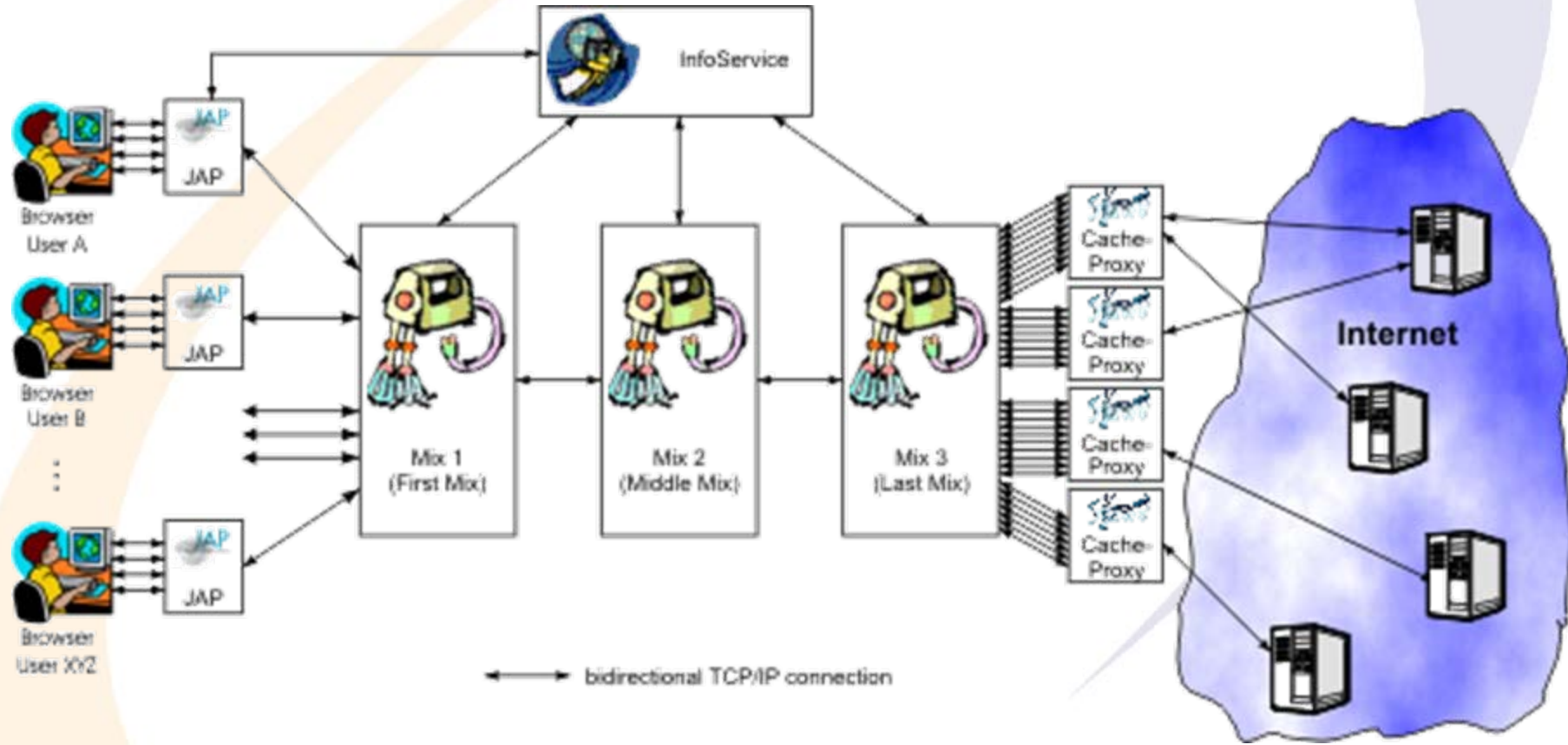
How Tor Works: 2

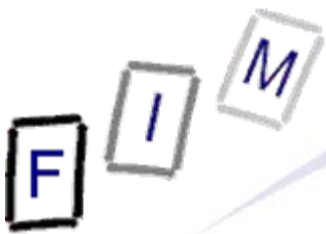
-  Tor node
-  unencrypted link
-  encrypted link





- 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
 - But how do you find them and know, whether you can trust 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**





Commercial proxy servers

- Commercial services exist, which provide you a proxy somewhere else (or allow selecting from a list)
 - Typically only a single computer at any time (no cascading)
 - The computer might be your own (no mixing with other users)
 - » Depends on service and payment; could also be shared (but you don't know with how many other users!)
- Uses:
 - Criminals
 - Accessing services with IP restrictions ("only within USA")
 - Anonymisation
- Price: ≈US\$ 10-20/month
- Several providers exist
- Some have been known to log data based on court orders
 - But not otherwise – at least not known!



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
 - E-Mail content can render the anonymisation meaningless
 - » "Send products to ...", signatures, metadata in attached files etc.



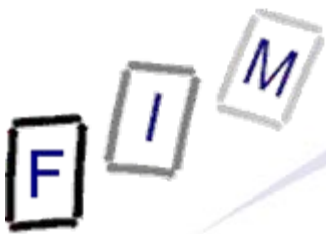
E-Mail anonymisation: Replies

- Depending on the system, answers might be possible
 - Some systems: Reference lists (Sender \Leftrightarrow pseudonym)
 - » But these are then in danger of break-ins or official searches!
 - Staged encryption
 - » Sender encrypts it three times, each forwarder "removes" one level of encryption
 - » Works for replies as well: Each stage adds one encryption layer
 - » Problem: Last hop still must know the original sender
 - But: This need not be the same computer who was sending the original message this is the reply to
 - No intermediate server knows the final destination E-Mail address
 - There is no association between message and its reply
 - » Still: Controlling the "exit" node allows some information leaks
 - Note: The content may still be encrypted!
 - But: "Forward" messages will usually/often be unencrypted!



E-Mail anonymisation: One-use accounts

- Simple E-Mail "anonymisation"
 - Create a new free account by an online service
 - Send the E-Mail
 - Receive some E-Mails
 - Delete the account or just ignore it and delete access data
- Problems:
 - Your IP address might still be present in the E-Mail
 - » Or stored by the provider!
 - Web bugs, ... in E-Mail might allow the sender to identify you
 - » Solution: No HTML, but text-only showing; but this might be difficult with a web-based E-Mail provider...
 - So not really anonymous – but still good against spam!
- Special providers exist which take care of the problems
 - But again: You have to trust them!



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
 - » Even 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
 - » Extreme cooling necessary; more a theoretical attack!
- DVDs, CD-ROMs, tapes: Shredding is the best method
- Note: Usually it is still detectable, **that** a drive was wiped!

Secure deletion of data



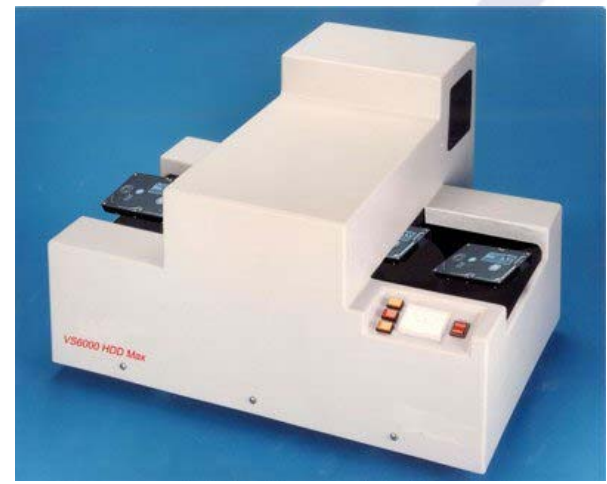
<http://www.flickr.com/photos/gmccarroll/341892350/in/set-72157594453290733/>



<http://www.ontrack.at/degausser/>



<http://www.periphman.com/degaussing/degaussers/PD8400.shtml>

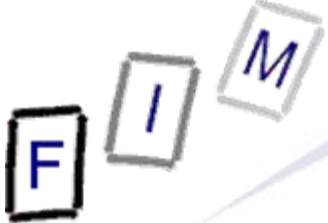


<http://www.degausser.co.uk/degauss/6000.htm>



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
 - » New article: Actually impossible (except very old/floppy disks)
- Different approaches to wiping exist:
 - » Attention: "All bits" need not be the same on physical surface!
 - Single pass: Random data, all zeros, or all ones
 - Triple pass: All Zeros, all ones, 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"
 - RLL: Encoding used ≈ 20 years ago on HDD; today on CDs/DVDs



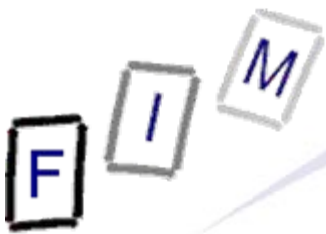
Destroying other media

- CD/DVD/Blue-Ray:
 - Cut into them with scissors
 - Shredding (many paper shredders have a CD/credit card slot)
 - Drill one or several holes into them (several at once)
- Magnetic tapes:
 - Remove tape from enclosure and cut into pieces
 - Drill into it from the side (through all layers of spindle)
- Practical problems: Mechanical dangers
 - Breaking plastic produces splinters
 - Drilling : Don't hold with hands, tapes might block the drill through wrapping around
- Practical problem: No difficulty for one or two media, but what about 50/100/500?

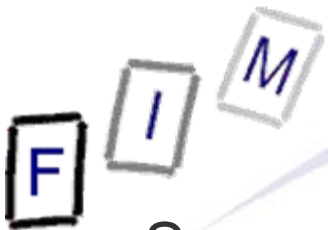


Selecting the correct privacy level

- 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 1 pass
 - » 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 off press, repressive countries, ...
 - Special care needed to be really anonymous (beside anon-proxy!)
 - Web anonymisation might be useful in rare cases
 - » Difficulty not to forget it: A single time without → No anonymity!
 - Various plugins, cookies, JavaScript can reveal the local IP as well!
 - » 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, where data retention doesn't apply
 - » The EU directive applies to the EU only ...
 - "Bullet proof" web hosting/ISP
 - "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 requires inspecting content (deep packet inspection!)
 - » Typically illegal (unless: police, secret service, ...)
 - » 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 **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
 - Typically used for Skype!



Conclusions

- 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!