



Data Retention

**Security and Privacy
Budapest 2009**

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>



- What is data retention and why is it necessary?
- The EU directive on data retention
 - What is retained and what not
 - Who is obliged
 - Who may access the data for what reasons
 - Safeguards: Security & privacy
 - Overview on national implementations: Austria, Germany
- Discussion of the directive: Aim, pros, cons
- Options for technical implementation of the directive
 - IP addresses
 - E-Mail communication
 - VoIP communication
- Alternatives



What is "data retention"?

- Data retention (DR) is the keeping of data for further use, which would have otherwise been deleted
 - Here, we are talking about "telecommunications DR"
 - Even more specific, about data retention of Internet comm.
- Subject of DR discussed here:
 - IP addresses
 - Communication acts within the Internet
 - » TCP connections, E-Mails, web sites visited, chat sessions etc.
- DR is nothing new and has existed for many decades
 - Pursuant to court orders telephones were fitted with tape recorders to identify the numbers dialled and all sound
- Problematic and currently hotly discussed is DR, which is independent of any suspicion:

Mandatory retention of all communication of all customers



Why is it needed?

- Basic idea: Going back in time!
 - DR allows investigating communication after it took place
 - » Typical "normal" DR only works from a point on ("realtime")
- Commercial companies employ data retention to learn about their customers
 - Examples: Google, Amazon
 - Typical usage: Personalization, invoicing, legal obligations, ...
- DR as discussed here:
 - Judicial proceedings (criminal and civil)
 - » File sharing, libel, hacking, espionage, ...
 - Police investigations
 - » Confirming suspicions, identifying accomplices, ...
 - Combating terrorism
 - » Uncovering terror networks, identifying accomplices



Computer Forensics and data retention

- Often a forensic examination only results in IP addresses
 - Examples: Tracing the origin of an E-Mail, intrusions
- As these occurred in the past, data from that point in time is required to identify the computer involved
 - Note: Dynamic IP addresses are allocated frequently to different persons, so they change over time!
 - Note: Typically not the computer but only the Internet connection can be identified, much less the actual user!
- If the retention occurs on the device under investigation (=log files), this provides additional information
- Another aspect of CF, e.g. after intrusions, is checking whether any kind of "malicious DR" took place
 - Keyloggers, snapshots, screenshots etc.



The EU directive

Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC

- Enacted: 15.3.2006; to be transposed: 15.9.2007
 - Internet part may be postponed up to 15.3.2009
 - » Many countries did this!
- Basic idea: Combat terrorism or "serious crimes"
 - ... investigation, detection and prosecution of serious crime, as defined by each Member State ...
- Problems: Proceedings whether it was enacted correctly
 - Directive would "disappear" if procedure/basis was incorrect
 - This would **not** affect the national laws!
 - Very similar (or the same) directive would almost assuredly be enacted again in very short time
 - » Might be even more strict!
- Background: Madrid bombings



What is to be retained?

- The following data should be collected:
 - » To identify both natural persons and legal entities
 - Trace and identify the source of a communication
 - » Calling telephone number, name and address of user, UserID
 - Trace and identify the destination of a communication
 - » Number dialled, final destination number (call forwarding, call transfers, ...), name and address, UserID
 - Identify the date, time, and duration of a communication
 - » Date & time of:
 - Start and end of communication of fixed network&mobile telephony
 - Log-in and log-off of the Internet access, IP address, UserID
 - Log-in and log-off of Internet E-Mail services
 - Log-in and log-off of Internet telephony services
 - Identify the type of communication
 - » Telephone service used (voice call, voicemail, fax, S/M/EMS, ...)
 - » Internet E-Mail and telephony: the Internet service used



What is to be retained?

- The following data should be collected:
 - Identify the communication equipment
 - » Fixed network telephony: Calling and called telephone numbers
 - » Mobile network telephony: Calling and called telephone numbers, IMSI and IMEI of caller and called
 - » Prepaid anon. services: Date, time and CellID of initial activation
 - » Internet access/E-Mail/telephony: Calling telephone number (modem dial-up), DSL/other endpoint of the communication
 - Identify the location of mobile communication equipment
 - » CellID of the start of the communication
 - » Geographic location of cells by CellID
- Period of retention: Minimum 6 month
 - But see e.g. Poland: Initially planned 15 year storage and finally settled on two years!

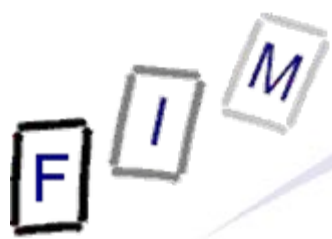


What is **not** to be retained?

- Unconnected calls
 - Calls, where the destination number does not exist
 - When the recipient doesn't answer this must be retained for the full time, but only if the information is already stored
 - » But there is no obligation to store it!
- Any content data (expressly forbidden)
 - This might be difficult in practice
 - » Example: Mails to order@sadomaso.com, help@drugabuse.com



- Providers of publicly available electronic communication services or of public communications networks
 - Only within the EU (i.e. within each member state)
- "Publicly available el. comm. services" =
 - Wholly/mainly conveyance of signals
 - available to general public
 - normally provided for remuneration
- "Public communications network" =
 - Electronic communications network
 - used wholly or mainly for the provision of
 - publicly available electronic communications services
 - » Explanation of the German law-draft: This excludes company-internal networks, PBX, E-Mail servers of universities providing services exclusively to students and faculty, and the communication infrastructure in the medical area



- Internet E-Mail & telephony: Obligations may apply only to data from the providers own services
 - » Stated in the (non-binding!) reasons
 - I.e., providers will probably not be required to log all traffic to port 25 on other servers, but only to their own server!
 - » Using the provider's SMTP server → DR
 - » Using the provider's IMAP/POP server → DR
 - » Sending an E-Mail directly to a server outside the EU would not be logged at all!
 - See e.g. the German law-draft!
 - » Using an IMAP/POP/webmail server outside the EU → No DR
 - » Only the IP address can be associated to the user
 - » Everything else would be **MOST** complex and expensive!
 - No monitoring port 25
 - No investigation of web traffic whether it is a webmail page



"Normally provided for remuneration"?

- Must be some kind of commercial undertaking
 - "Private" use is excluded
 - » Does this apply to Fon/Freifunk?
- Remuneration must not stem from the customer
 - Commercial free TV: Paid for by the advertisers
- Unclear: Whether "normally" depends on this service provider or the service in general
 - Always free at this provider, but such a service must be paid at all other providers?
 - » Public WLAN hotspots – You have to pay for at hotels, ISPs, ...?
 - This type of service is free, but "better" ones cost money?
 - » Typical example of webmail: Basic version is free, but spam filtering, more space, etc. costs money!



"Public service"?

- What restrictions are necessary to be "non-public"?
- One approach: In effect the access depends "only" on the **payment** for the service
 - Universities would be excluded; in Austria students must have the "Matura" to be accepted (but then **must** be!)
 - Companies would be excluded, as employees will only be hired if the company needs them and finds the individual "ok"
 - Problem: ISPs!
 - » They accept customers only in the area they can actually supply their service (→ characteristic of the provider!)
 - » But this could be seen as a "local" public service
 - » Also, this depends on the characteristic of the service itself
- Counterargument: Set up company for commercially providing service only to a clearly defined subset of citizens!
 - E.g. some car insurances are available solely to women!
 - Internet access only with computer driving license ?!?



Example: Universities

- Position of the Austrian ministry of education:
Universities are not public → No DR (would be too costly!)
 - "Club of all persons allowed to study" → Only members can obtain Internet access from an associated company
 - » Is this still "public?"
 - Universities with access restrictions?
 - » E.g., universities of applied sciences (Fachhochschulen)
- But: "normally provided for remuneration"
 - Austria studies are "free", as the study costs are only nominal
 - This does not apply to fully-paid courses
 - » Which also exist at many universities!
 - Does this affect the university not at all, only these studies, or is the whole university then suddenly "for remuneration"?
 - Consider also private universities!



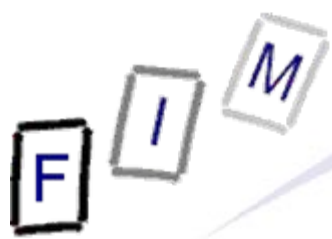
Who may access the data?

- Who may access the data?
 - Only "competent national authorities"
 - » Will be defined by each member state!
 - Only in specific cases
 - » Not to be used for general computerized searches
 - In accordance with national law
- National laws for procedures and conditions must:
 - adhere to the necessity and proportionality principle
 - conform to European law, national law, and especially the European convention on human rights (ECHR)
- What is missing?
 - What may be done with the data afterwards?
 - » Indefinite usage/storage?
 - » Or do the "normal" privacy rules apply? ← Presumably!



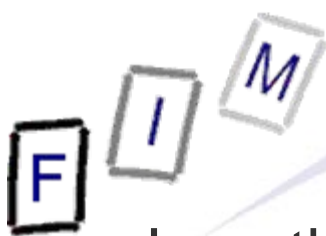
Privacy safeguards through providers

- The data collected must be treated according to the normal privacy laws/directive unless changed for DR
- Additional explicit requirement:
 - Access may only be possible to personnel specifically authorized to do so
 - » Additional encryption or authorization (log-in), ... necessary
 - Data must be destroyed at the end of the retention period
 - » Unless it was accessed and preserved
 - E.g. for ongoing proceedings
 - » This is technically not that easy to realize!
 - Is this to be done daily/weekly/monthly/yearly?
 - How to exclude this specific data from deletion?
 - » Unclear is the collision with other rules/permissions:
What if this data is necessary for other legal purposes and might be stored, used, ... according to privacy laws?



Security safeguards through providers

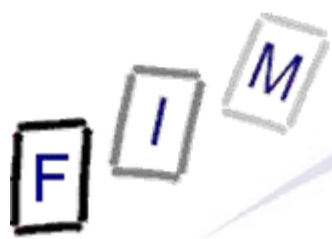
- Data must be of the same quality, and subject to the same protection, as the data on the network
 - Quality: States that we may not "reduce" the data in any way
 - Protection: If we don't secure the data on the IP network at all, do we have to protect the stored IP addresses at all?
 - » But see next requirement!
- Data must be subject to appropriate technical and organisational measures to protect it against
 - » accidental or unlawful destruction,
 - » accidental loss or alteration, or
 - » unauthorized or unlawful storage, processing, access or disclosure.
- This is technically not that easy and requires probably extensive precautions (to be further detailed in laws)



Data retention in Austria

- Length of draft: 8 pages, including explanations!
- Data will be stored for 6 month
 - The exact calculation (start) of the period is unclear!
 - » Wording is bad: All data would have to be deleted, even name and address of the customers!
 - Protocol of access to the data may only contain the category of the data accessed, but not the data itself
 - » Because the protocol itself may not be deleted after 6 month!
 - » Protocol does not contain, where the data was sent to
 - Only the reference to the court decision authorizing it
- Applicability ("serious crime") is very wide:
 - All crimes with more than 1 year prison term
 - » Includes even carelessness "crimes"!
 - » But this does NOT include non-commercial copyright infractions
 - This would even be a reduction compared to the current situation:
At the moment the data can be accessed via courts if present

Attention: This is according to the draft! The law has not yet been enacted and may yet be modified!



Data retention in Austria

- It is unclear, who may access the data and for what reasons
 - Only criminal procedure or also civil/administrative ones?
 - » Probably for everything, but only through courts
- No compensation for the ISPs
 - "There are no additional costs, as only data is to be retained, which is already yet stored for invoicing" ???
- Problematic is especially that there are no exceptions at all
 - Currently several special cases exist, where e.g. telephone interception is forbidden or restricted
 - » Examples: Lawyers, doctors, ...
 - See also the "whistleblowing" hotlines required by US law
 - » These would no longer be anonymous at all ...
- Storing the final destination of telephone calls is allegedly impossible in some cases
 - E.g. forwarding to a different provider



Data retention in Germany

- Length of draft: 193 pages, including explanations!
 - But includes other surveillance measures (DR: >19 pages)
- Retention period: 6 month
 - Must be stored within the EU
 - Deletion within one month after the 6 month period elapsed
- On call-transfers **every step** must be logged
- Data may be accessed only by certain institutions
 - Must be enabled for each area within the law
 - » Example: § 100g StPO references §113a TKG (=DR)
 - May only happen on single cases
 - » No "general" access, e.g. all data from a certain area
- If the telecommunication service is not provided directly, the provider must ensure that someone else retains the data
 - But this is to be interpreted extensively; e.g. call forwarding



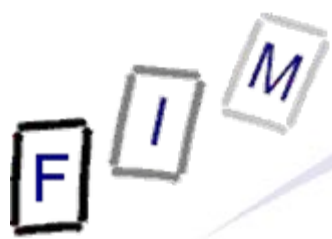
Data retention in Germany

- Data may be used for:
 - All criminal proceedings
 - » Must be explicitly provided for in law
 - Included: All crimes committed through telecommunication, serious crimes which are "also in the specific instance serious"
 - This includes all copyright infractions in the Internet ("through telecommunication")!
 - » Note: Filesharing in Germany is now usually handled like this:
 - Media company starts criminal proceedings
 - Public attorney identifies the person
 - » Newer decisions: This is not done any more → No identification anymore!
 - Public attorney drops the criminal proceedings because of little guilt
 - Media company inspects the files
 - Media company starts private proceedings
 - To prevent significant dangers for public security
 - Constitutional protection of country and states, secret service, and military secret service



Data retention in Germany

- Anonymisation services must log all the anonymisations!
 - Law: Who changes some data must log **when** it was changed and **what** was changed to **what**
 - All anonymisation services within Germany would be practically "abolished"
 - » Please note: There is a law requiring telephony providers to offer anonymous services if possible (not very strictly enforced!)
- Only data which is created or processed must be retained
 - I.e., mere transmission is not affected!
 - This means e.g. for E-Mail: Only source and destination must store the addresses/IDs/...
- Data stored **only** because of DR may not be used for anything else
 - Example: No analysis for marketing purposes!



What are the aims of the directive

- "Prevention, investigation, detection, and prosecution of criminal offences"
 - Prevention is mentioned only at the beginning
- Making sure that the anonymity in the Internet is not used to create a lawless area in effect
 - That laws **do** apply is clear nowadays
 - But currently they cannot be enforced in many cases, as the perpetrators are completely anonymous
 - » Only IP address known → "untraceable" to a single person
 - » Shipping to a physical address is also no sure identification
 - Similar to license plates on cars!
- Help in identifying the "network"/the criminal **after the fact**
 - If known in advance → current wiretapping, search etc. possibilities are already sufficient!



Arguments for the directive

- Recent terror attacks could be traced back to the terrorists or some accomplices through their mobile phone calls
 - But this was only possible, because this data was available!
 - » And with pre-paid phones, flat-rates, etc. this is less likely
 - Other targets are organized crime, phishing, fraud, child pornography/misuse, etc.
- While e.g. P2P filesharing by students is not really a "serious crime", it is still illegal
 - If no tracing is possible, copyright is essentially abolished in the Internet!
- Some countries do already have DR
 - Different models in various countries are problematic for transnational providers (especially mobile phones)



Past regulations: The Convention on Cybercrime

- An international conventions to combat cybercrime
 - Several years old (23.11.2001)
 - But in many countries not yet transposed to law
- This international convention included e.g. provisions:
 - "Quick-freeze" (Art. 16): Expeditious preservation of specified computer data, **including traffic data!**
 - » Preservation and maintenance for up to 90 days to allow for disclosure (e.g. to go through an judicial approval process)
 - Partial disclosure: To enable tracing the traffic to other providers to order a quick-freeze there
 - Secret real-time collection of traffic data
 - Secret interception of content data



- It is very easy to avoid the data retention
 - Use pre-paid mobile phones, EU-external webmail/servers, encryption, Internet cafes, anonymisation services, ...
 - No serious criminal is likely to be caught, unless he is very careless or makes outrageous errors
 - » Might even increase the use of such services!
 - So whether it can actually reach its aims is very doubtful!
 - Therefore unsuitable as deterrent
- Storing mobile phone locations allows interesting possibilities
 - E.g. in divorce proceedings, but also as alibis
 - » This has in general little to do with serious crimes!
 - However: Presence of phone and even communication does not necessarily mean, that the **owner** was really there,...



- Communication analysis is possible
 - If a person repeatedly calls a psychiatrist during his office hours, this hints at an illness, i.e. sensitive data!
 - Analysis of who talks with whom is not really restricted
 - » There is no general access, yes, but still ...
 - But if a single person is known, then a network can be traced from this person on to all others
 - » This is the intention: Combating terrorism
 - » But when data exists, it can be used for other things as well!
 - Example Germany: Access for secret service!
- Unification doesn't really take place:
 - Every country can have:
 - » Different duration
 - » Different access procedures (and different entitled institutions)
 - » Different crimes (unimportant for the providers)

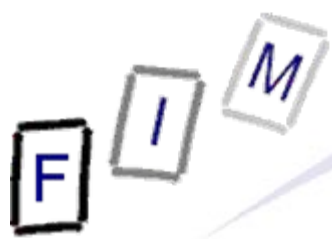


Problems of the directive (3)

- If you are a private person and offer a public service (which is normally offered for remuneration), DR applies
 - Example: Public WLAN hotspot → DR might be necessary
 - » This might include the obligation to identify all users!
 - » Therefore probably a problem for local public hotspots too!
 - Depends on "remuneration" → This provider or generally?
 - This also applies to E-Mail, but not to webhosting, chat, discussion groups, NetNews, ..
- Not only criminals are monitored, but everyone
 - See George Orwell: 1984!
 - Everyone is a suspect per definition
 - » And then might have to prove his/her innocence!
- If data exists, it will be used
 - The catalogue of crimes will continuously be expanded
 - » Public outrage → "Don't let them get away with it!"



- Data is perhaps not always stored very securely
 - Not in the interest of the provider!
 - Hacking of the server or unauthorized access would lead to enormous personal information!
- In the end, all the customers will have to pay for it
 - » Some estimates: 10-15 % price increase, end of business for small providers and some webmail providers
 - Large ones might move outside the EU
 - The companies will not "swallow" this from their profits
 - Even when paid for by the state → Taxes!
- Might be against the constitution
 - Freedom of communication, privacy, ...
- A more pressing problem seems to be accessing existing foreign data, which is extremely slow or impossible



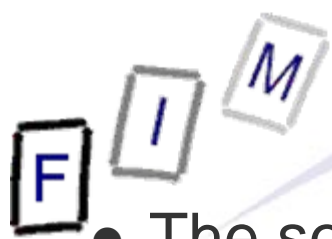
Study by the German BKA (15.11.2005)

- Study by the German Federal Policy: The solving ratio of crimes would at most be increased through DR by 0,006 %!
 - 381 (=0,006%) cases could not (?) be solved because of missing communication data in several years
 - » Two of them were from organized crime/terrorism
 - » 36% were fraud and computer fraud
 - » Not all of them might have been solved with data!
 - They just could not investigate further as no data was available
- Currently the ratio of cases solved is in telecommunications higher, and in internet fraud and software piracy very much higher than the average ratio
- Note: This study must be seen as suspect!
 - There is no mention what cases (should) have been reported!
 - » Or whether it was obligatory to answer



Technical implementation: IP address

- An ISP must store whenever a customer is connected to the Internet, i.e. dials in, powers his router, ...
 - Trivial with static IP addresses; these are stored anyway
 - Dynamic IP addresses: DR must take place now
 - » Currently: Typically already in place for accounting
 - No accounting allowed (privacy!) for customers with fully unlimited data transfer (amount and time; "fair use" does require it!)
- Technically not that hard to implement, but the storage, backup, access-restrictions, logging, etc. will require new software and regular maintenance



Technical implementation: E-Mail communication

- The source must store all the data
 - Example: An ISP providing an SMTP server as a proxy must store who sent an E-Mail at what time to which recipient
 - » Users send them directly without an SMTP server: No DR!
- The destination must store all the data
 - Example: An ISP receiving an E-Mail for a customer must store from whom and to whom it was sent
 - » Sending to a host under end-user control: No DR!
- The access/receipt must be logged
 - Example: When a user accesses his E-Mail by POP or IMAP, or logs in to a webmail service the ISP must log this access
 - Note: Regular checking (e.g. all 5 min.) is common → Huge log!
 - » If the service is outside the EU: No DR!
- Implementation requires server modifications
 - Logging currently possible, but not necessarily a single line/in a DB; may contain other data, e.g. the subject, IDs, ...



Technical implementation: VoIP communication

- The provider of the telephony service must store who called whom at what time
 - Example: Skype must store each and every connection
 - » Skype-IDs and IP addresses
 - If Skype is located outside of the EU:
 - » If no service is offered into the EU, there is no DR obligation
 - If some users employ it "inofficially" there might still be no DR
 - » Problem: Skype cannot easily locate its users
 - What about US customers travelling within the EU?
 - Not accepting any connection from an IP address within the EU?
 - » Just "drop" EU then? Skype probably not, but smaller ones ...
 - Important question: Is Skype an "Internet telephony service"?
- Many such services are for free in large areas
 - This data is probably currently not stored at all!
 - Only the parts to be paid for!
- Note: ICQ (=text chat) is **not** internet telephony!



- Secret online surveillance
 - Would yield even more information
 - But mostly only usable "forward"
 - » Only stored mails can be investigated, but not deleted ones
 - These might be recoverable by computer forensic, but this is probably too complicated to add to online surveillance software!
- Audio-/Video surveillance
 - When monitoring the room with the computer, much information, e.g. internet telephony, can be gathered as well
 - » E-Mails are probably rather difficult to monitor
 - Usable only forward
 - Modifications of the data by the investigator impossible!
- Quick-freeze
 - Usable only forward



- Some measure of DR is probably necessary
 - To avoid the Internet becoming completely anonymous
 - » I.e. retaining solely the IP address for some time
- Logging individual communication acts is not necessary
 - E-Mail, location of mobile phones, telephone calls
 - Reasons:
 - » Too easy to subvert
 - » Not worth the effort: Very limited results
 - » Amassing data which will only be used extremely rarely
 - Or completely automatic, which is even more frightening!
- Data retention as presented here will come
 - If the directive is declared illegal by the court, a replacement will be created
 - National laws will be enacted anyway

F I M

Questions?

Thank you for your attention!