Legal issues in measurements

Markus Peuhkuri

2006-04-20

Lecture topics

- Legal issues governing measurements
  - operator networks
  - end user organisations
- Focus on Finland

After this lecture you should

- Know how to make measurements and not to end up to headlines
- Know key legal resources
- Have some knowledge to challenge lawyer's immediate NO

What is the problem

- User data sensitive
  - a private message is confidential by Finnish constitution

    > The secrecy of correspondence, telephony and other confidential communications is inviolable.
- Protocol data sensitive
  - protocol fields may carry *identification* information

Concepts of legal system

Acts * are given by Parliament	laki
Decrees * are given by Ministries	asetus
Regulations * are given by officials to whom right is given by an Act or a Decree	määräys
Special enactment * dictates ruling different from general act* in a specific situation	erityislaki yleislaki

(Data) security governance in Finland

- Ministry of Transport and Communications* — Liikenne- ja viestintäministeriö
  - FICORA (Finnish Communications Regulatory Authority*) — Viestintävirasto
- Ministry of Justice* — Oikeusministeriö
  - Office of the Data Protection Ombudsman* — Tietosuojavaltuutetun toimisto
- Ministry of Trade and Industry* — Kauppa- ja teollisuusministeriö
  - Consumer Agency* (Consumer Ombudsman*) — Kuluttajavirasto
Kuluttaja-asiamies

  - National Emergency Supply Agency*
- Ministry of the Interior*
  - Police

Huoltovarmuuskesk
Sisäministeriö

Key acts

- Personal Data Act* (523/1999) — Henkilötietolaki
- Act on the Protection of Privacy in Electronic Communication* (516/2004) — Sähköisen viestinnän tietosuoja-laki
- Communications Market Act* (393/2003) — Viestintämarkkinala
- Act on the Protection of Privacy in Working Life* (759/2004) — Laki yksityi-syyden suo-jasta työelä-mässä

Personal Data Act

- General act on processing of personal data
- Furthermore 650 acts gives detailed instructions
- Key terms

personal data * information on a private individual related to an identifiable person or family — henkilötieto

processing of personal data * is any action done on personal data — henkilötietojen käsittely

personal data file * is a storage where personal data can be retrieved easily and at reasonable cost — henkilörekisteri

controller * who determine use of data file — rekisterinpitäjä

data subject * is subject of personal data — rekisteröity

- Duty of care
  - good processing practice
  - safeguards for private information
- Use of personal data must have a defined purpose that is a real one and not one dictated by technology
- Data may not be used for a purpose that is incompatible with original purpose
  - historical, scientific and statistical purposes are not incompatible

Act on the Protection of Privacy in Electronic Communication* (516/2004)

sähköisen viestinnän tietosuo-jalaki

- Replaces Act on the Protection of Privacy and Data Security in Telecommunications 22.4.1999/565
- Implements EC Directive on Privacy and Electronic Communications* (2002/58/EC) — SVTSL
- Definitions — sähköisen viestinnän tietosuojadi-rektiivi

message * is a phone call, e-mail message, SMS message, voice message or any comparable message sent in — viesti

communications network * is any system using electromagnetic means to transport message — viestintäverkko

public communications network * is a network available to set of users without any prior restriction — julkinen vv

telecommunications operator network- or service provider

network service provision of a communications network by a telecommunications operator for providing

communications service means the transmission, distribution or provision of messages

value added service using identification data or location

identification data associated to subscriber or user

location data indicates the geographic location

subscriber a legal person or a natural person

corporate or association subscriber

user a natural person

information security administrative and technical measures to protect data

processing means collecting, saving, organising, using, transferring, disclosing, storing, modifying, combining, protecting, removing, destroying and other similar actions.

- Covers
  - public communication networks
  - networks attached to public networks
  - secrecy and privacy in internal (restricted) networks

Act on the Protection of Privacy

- Sets demand on
  - network and service providers
  - value-add service providers
  - corporate subscribers
  - users of network
- Handling of *identification data*
  - any data that records existence or details of a message
- Corporate subscriber
  - organisation, that has users using services provided
  - may also be the other party in communications
  - usually a bystander
  - ultimately responsible even if services outsourced

Who has a right to handle identification data

- To realise services
  - even automatic handling for relaying is handling
- To implement data security
  - firewalls, virus scanners
  - must not infer with legal communication
- For charging
  - in most cases, no reason to reveal B-number
    $\Rightarrow$ aggregate information sufficient
- To improve technical implementation
  - only aggregate or anonymous information
  - includes also statistical, scientific use

- To resolve technical problems
- To resolve misuse
  - *not* to follow where a employee visits or what messages sends (unless identified as virus)
  - misuse must have some direct costs
- Communicating parities
- If permission by one of communicating parties

How to handle identification data

- Only when needed
- Only as much as needed
- Only those whose duties it belongs to
- Handing information over only to those that have right
- Service provider must have audit trail for two years
- Professional discretion must be maintained

Information security and privacy

- Corporate subscriber *must* take case of identification data security
- Threats on information security
  - may take actions to protect system security
  - remove malicious payload
  - refuse from accepting messages
- Must not exaggerate actions
  - no limit freedom of speech or privacy
  - must stop as soon as there is no immediate need
  - filtering should be done without accessing message content

Act on the Protection of Privacy in Working Life

- A special act for Personal Data Act and Act on the Protection of Privacy in Electronic Communication
- Rules for
  - handling employee personal data
  - tests for employees
  - technical surveillance
  - opening emails
- Strict rules for what is allowed
  - uneven situation between employer and employee: “this is ok, isn’t it — or do you want to start looking for a new job”
- Technical supervising and data networks use
  - employees must be informed in cooperation procedures

How to measure, then

- Get rid of identification information: once data does not contain
  - identification data
- It is not anymore
  - personal data
  - telecommunications identification data
- And thus it does not form a
  - personal data file
- No user data should be captured

Should users be informed

- In corporation, yes
  - part of cooperation discussions / consulting with general trustee* pääluottamusmies
  - should include what is measured
- In public networks, no
  - telecommunications provider has right[1] to develop one's systems
  - also long-term development

When IP address is an identification information

- If it identifies a person or a household
- Thus, it usually is not when it is
  - server IP address
  - dynamically allocated. Current consensus within IT community is that if addresses are allocated using DHCP protocol [1] they are not identification information. However, I would not try to test that on court. Remember that in normal course of DHCP operation a host will maintain the same IP address indefinite time, even across reboots.
  - some of technical multicast addresses
- How one can tell the difference

Removing sensitive information

- Address anonymisation
  - refer to previous lecture
- One may end with semi-sensitive data
  - accidental disclosure avoided
  - /24 prefixes mostly OK
- Organisational data may be sensitive
  - lots of traffic from organisation O to questionable sites S (refer to previous lecture about prefix-preserving anonymisation)
  - questionable traffic

[1] Actually, an obligation.

Problems in IP address anonymisation

- There are a finite number of users
- Taken traffic of random IP address, it is usually not possible to determine whose traffic it is
- However, it is often possible to answer opposite
  - is this IP address person X
  - if something *a priori* is known about traffic by X
- IPv6 provides more addresses
  - possible to change IP addresses
  - simultaneous use of multiple IP addresses
  - number of human users it does not increase

Data example: Traffic captured on open public WLAN AP on café

- Addresses dynamically allocated, remember to take care of MAC addresses if applicable
  $\Rightarrow$ client IP addresses are not identifiers
- Peer addresses may identify users, for example using home email server would identify user
- Top-sites do not identify users
- Group peer addresses into two groups
  - popular sites used by multiple customers
  - private sites used by few users
- Anonymise private site addresses

Data example: Traffic captured from ISP core network

- Possibly no information about netblocks
  - an ISP should have that information
  - home, corporate users $\Leftrightarrow$ servers
  - large number of addresses
- Any address possibly an identifying information
  $\Rightarrow$ safest to anonymise all
- Routing information should be preserved

Data example: Analysing corporate data network

- Interesting questions
  - how much total traffic
  - which applications consume most of bandwidth
  - response times of servers
- Close monitoring of communication not appropriate
  - e.g. sites visited, emails sent
- For full packet capture, client addresses should be anonymised
- Server addresses may stay intact
- Note different rules for troubleshooting acute technical problems

Traffic data life cycle

- Each data has a life time, possibly an indefinite one
  - data is generated
  - data is used actively
  - data is archived
  - data is destroyed
- Data sensitivity may change over time
- Must have a plan to take care of whole life cycle

Access to data

- If data contains identification information, its use must be monitored
  - enumerate who has access to that data based one's tasks
  - what searches have been made: auditing system
- Desensitised data may also need controls
  - enumerate who has access to that data based one's tasks
- Remember appropriate professional discretion agreements

Handling results

- Not to disclose too much information
- Aggregate summaries are usually safe
- Distribution of results

How about rest of the world

- Privacy laws usually much more lax
  - employer has more rights
- Old laws
- EU directive should lead to similar laws
- Multi-jurisdiction operations can be problematic

Conclusion

- Reasonable measurements are possible
- Take special care on handling of identification information

References

[1] R. Droms. Dynamic Host Configuration Protocol. Request for Comments RFC 2131, Internet Engineering Task Force, March 1997. (Internet Draft Standard) (Updated by RFC3396) (Obsoletes RFC1541). URL:`http://www.ietf.org/rfc/rfc2131.txt`.