Privacy, Data Protection Law and Flow Data Anonymisation: requirements, issues, and challenges

Elisa Boschi, Hitachi Europe
Ralph Gramigna, KPMG

Acknowledgement: M. Bossardt (KPMG), D. Battisti (ETH)

Outline

- Review of law principles and requirements on data protection
  - European viewpoint
  - What is personal data?
  - Why is data protection law relevant for network monitoring?
  - Law principles overview
- The role of flow data anonymisation to support data protection
  - Discussion on its applicability and weaknesses
  - Suggestions for future steps

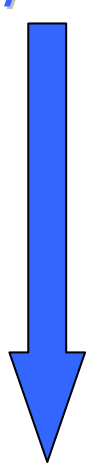
Data Protection Law: EU Directives

- Goal: protect the privacy of individuals
  - Not limited to information confidentiality
- EU Directives define the the minimum law requirements to be implemented by each EU member state
  - Applicable to international data transfers with EU
- Relevant to data protection:
  - Directive 1995/46/EC - on data protection
  - Directive 2002/58/EC - on privacy and electronic communications

Applicability and Personal Data

- Directive 95/46/EC applies to the

  „*processing* of *personal data*“



"*any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly* ***or indirectly****, in particular by reference to an identification number or to one or more factors specific to his ... identity*".

"any operation performed upon personal data, such as e.g. collection, storage, adaptation or alteration, consultation, disclosure by transmission, dissemination or otherwise making available, alignment or combination, erasure or destruction"

- Note: in some countries (e.g. Switzerland) this applies to „legal entities“ as well

Applicability to Network Monitoring

- *Indirect identification data comprise any information that may lead to identification of the data subject through association with other available information*
  - information available to the entity in charge of the data processing (ISP),
  - any information possessed by third parties
- IP addresses can identify someone “directly”
  - Esp. legal entities
- Many more attributes in a flow record can contribute to identifying someone “indirectly”

Principles: legitimation for processing

1. Consent
2. Data processing is „*necessary for the performance of a contract to which the data subject is a party*"
3. …

- Processing must be limited to specified purposes

- Further processing of data for historical, statistical or scientific purposes is possible provided that appropriate safeguards are provided
  - Left to national laws

Principles: Information of the Subject

The subject must be informed about:

1. Identity of the data controller
2. Purpose of the processing
3. Other information, e.g. the recipient of the data.

- It does not apply to scientific research, **IF** the provision of such information
  - proves impossible
  - would involve a disproportionate effort

- Appropriate safeguards must be provided
  - Their specification is let to national law

Border Crossing

- Transfer to third countries is generally possible if the third country ensures an adequate level of protection

http://ec.europa.eu/justice_home/fsj/privacy/thrid countries/index_en.htm

- E.g.
  - ✓ Switzerland, Canada, Argentina
  - ✗ USA (except Safe Harbor)

Traffic data and location data

- Introduced in Directive 2002/58/EC
  - *Traffic data: any data processed for the purpose of the conveyance of a communication or for the billing thereof*
  - *Location data: data indicating the geographic position of the terminal equipment of a user*

- Objectives:
  - Minimise the processing of personal data
  - Use anonymous or pseudonymous data where possible.

- „Anonymous“ = it is no longer possible to identify the data subject

Processing of Traffic and Location Data

- Traffic and location data relating to subscribers and users must be erased or made anonymous when no longer needed

- The processing of traffic data must be restricted
  - To persons acting under authority of providers
  - To certain activities (e.g. traffic management, fraud detection...)

- Location data can be processed only if
  - There is consent, or
  - Data is made anonymous

The Role of Flow Data Anonymisation to Support Data Protection

- The well known problem:
  - The more you anonymise the better privacy is protected...
  - ...but the less useful the data
- Anonymisation aims at removing sensitive information referring to an individual
- Attacks to anonymisation schemes have proved that those schemes could be broken allowing to "indirectly" identify people.
- Are known flow anonymisation techniques effective in protecting the privacy of individuals?

(4) Anonymization Techniques

Field to be anonymized:

IP address

IP	Truncation	Permutation	Black Marker	Prefix Preserving
135.98.111.17	135.98	141. 2. 32.37	10.1.1.1	22.131.88.67
135.98.111.128	135.98	41.12.96. 67	10.1.1.1	22.131.88.157
135.98.132.37	135.98	142.72.8.5	10.1.1.1	22.131.201.29
141.161.3.3	141.161	21.33.4.1	10.1.1.1	12.192.32.51
141.72.8.5	141.72	11.14.96.118	10.1.1.1	12.78.201.97
32.53.48.1	32.53	12.161.3.3	10.1.1.1	31.197.3.82

Some Anonymisation Attack Methods

- Data injection → injecting information to be logged with the purpose of later recognizing that data in the anonymized trace
- Fingerprinting → matching attributes of an anonymized object against those of a known object (e.g. web server) to discover a mapping between them
- Semantic attacks → system is exploited in a way that the victim thinks to do something, but he is doing something different. The attacker may infer part of the unanonymized IP address by exploiting the semantics of prefix preserving.
- Structure recognition → recognizing structure between anonymized and unanonymized objects

Attacks vs. Anonymisation Techniques

Anonymisation / Attacks	Prefix-preserving	Cryptographic approach	Truncation	Permutation
Semantic attack	■	■		
Cryptographic attack	■	■		
Data Injection	■		■	■
Fingerprinting	■		■	■
Structure Recognition	■		■	■

- ■ the attack can be used, (partial) results achieved

Conclusions

- We need to pay attention to data protection laws
- Anonymisation is part of the solution to protecting privacy, but
  - Research is still needed
  - This is not only a technical problem; a technical solution alone is not enough
- Legal solutions, policies, guidelines, interdisciplinary work are needed
- Anonymisation support is needed in standard flow data export protocols such as IPFIX