TECHNISCHE UNIVERSITÄT BERLIN

Bachelor's/Master's thesis

Revealing and structural analysis of personal data disclosure networks in real-world scenarios

Context

According to Art. 12 of the EU General Data Protection Regulation (GDPR) data controllers shall take "appropriate measures" to provide transparent information about (for instance) which personal data is collected by whom, for which purposes, under which legal bases etc. following the legal requirements of at least the Art. 13 and 14 GDPR. Besides privacy policies written in natural language, also electronic means in form of transparency frameworks [1] are used to fulfill this obligation.

State of the Art & Problem

In contrast to the underlying intention of the GDPR to reveal all personal data processing activities (incl. third country transfers or automated decision making), current privacy policies and electronic means (if at all available) do not reveal personal data disclosure networks across multiple data controllers or services, respectively. Several problems arise while trying to discover (potentially unwanted) indirect personal data sharing activities with regards to hiddenly related service providers or ambiguously stated transparency information. Also, technical means for a structural analysis of personal data disclosure networks are missing entirely. Thus, an automated evaluation for the risks of a data breach or trust enabling tools in this regard are not feasible [2].

Thesis Topic & Goal

Within this thesis, an extensible tool for exploring GDPR transparency information shall be designed, implemented and evaluated. Within this framework preferably a set of methods for social network analysis [3], data analysis, information retrieval etc. shall be used to reveal personal data sharing. Therefore, information about several real-world services (from said transparency frameworks and privacy policy analyses) need to be imported and represented in a structured way (e.g. through a graph database + schema). Afterwards, a selection of the aforementioned methods shall be applied. Eventually, the structural properties of then identified networks are to be compared (identifying sharing clusters, chains etc.). These properties tend to result in input parameters for a new metric to express the risks of a data breach. Ideally, the developed application should rely on web technologies. Ultimately, the tool is evaluated on its software qualities (e.g. performance) and the value added for data subjects or supervisory authorities. In general, the use of existing tools and prior work [4] to built upon is highly encouraged.

Contact: Elias Grünewald
eg@ise.tu-berlin.de

Skills

- Good (web) programming skills and
  a solid understanding of data analysis methods
- Interest in privacy/data protection, the GDPR, and knowledge discovery related topics

References

[1] https://iabeurope.eu/transparency-consent-framework
[2] Milne, G. R., Pettinico, G., Hajjat, F. M., & Markos, E. (2017). Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing. Journal of Consumer Affairs, 51(1), 133-161.
[3] Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge University Press.
[4] https://github.com/Transparency-Information-Language/transparency-analysis-platform

Our Mission:
Our lectures cover fundamental methods and techniques in the areas of service computing, cloud computing, and enterprise computing. We like to engage students in hands-on building of distributed information systems and to take an interdisciplinary approach to evaluating such systems. Through a close mentoring of students, especially in our seminars, we aim to introduce students to our ongoing research and to excite them to do future studies and research with us.

ISEngineering
Information Systems Engineering