Research question, hypothesis and preliminary design

Bartosz Czapski, TU060, ASD, D10123621

Domain and scope of research - ACM 2012

A: Social and professional topics → Computing / technology policy → Privacy policies
(Senarath & Arachchilage, 2019; Biega et al., 2020; Senarath & Arachchilage, 2018)

B: Social and professional topics → Computing / technology policy → Government technology policy → Governmental regulations
(Pernot-Leplay, 2020; Sullivan, 2019)

C: Security and privacy → Human and societal aspects of security and privacy → Social aspects of security and privacy
(Lee, 2020; Balapour et al., 2020; Li et al., 2019)

D: Security and privacy → Human and societal aspects of security and privacy → Privacy protections
(Amato et al., 2020; Mousavi et al., 2020; Mazel et al., 2019)

E: Security and privacy → Software and application security → Domain-specific security and privacy architectures
(Chen, 2020; Barbosa et al., 2019; Antignac, Le Métayer, 2014)

Domain: Data Governance and Privacy Implementation

Scope: Research on design and implementation of Data Minimization methodology during systems design process.

Gaps in the literature and research question

Data Minimization is a principle of collecting, storing and managing minimal amount of data required by the system to fulfil its designed purpose. Senarath and Arachchilage (2019) attempted to design Privacy Engineering Methodology (PEM) by conducting three studies: Study I with 24 participants, Study II - 9 participants, and Study III - 149 participants. Authors missed opportunity of involving multinational organizations where PEM could be tested among larger, and more organized group of participants. Using multiple, different scenarios instead of health application used by Senarath and Arachchilage (2018) could potentially provide broader, more complete data. Study group size was addressed in Balapour, et al. (2020) research where authors used 1544 participants divided into two groups, tasked with collection of more and less sensitive data. Authors used five-point Likert scale (from “strongly disagree” to “strongly agree”) which might be considered insufficient. Additionally, all participants of Balapour, et al. (2020) research were located in U.S.A. which considering how Data Privacy is regulated across different jurisdictions, might have affected their responses (Pernot-Leplay, 2020). Implementation of Vulnerability-Privacy Concern-Resistance (VPR) framework in Lee’s (2020) research focused on Internet of Things (IoT) where author’s goal was to determine factors that can inhibit acceptance of the technology among participants. Lee (2020) examined response from 265 participants, where 66% had previous IoT home experience. Increase in number of participants without IoT experience, and more balanced group could greatly improve proposed model. Barbosa, et al. (2019) proposed Privacy by Evidence (PbE) methodology for implementing privacy guidelines into systems development process. Case studies used in this research are very generic, and missing details on conditions and research concluded. Additionally, use of weights and calculation of the final metrics is unclear and oversimplified. As mentioned by Barbosa, et al. (2019) “, a quantitative experiment to compare results between team of developers using PbE and team of developers not following methodology would be a great addition to the results of experiment.”

Research Question: Can early adoption of Data Minimization methodology during system design improve data privacy?

Hypothesis

Null hypothesis H_0:

Early implementation of data minimisation methodology has nugatory effect on data collection during systems design.

Alternative hypothesis H_1:

Early implementation of data minimization methodology significantly improves data collection process and data privacy during systems design.

Feasibility of the study

TASK	DESCRIPTION	TIMEFRAME
Corporate Engagement	Involvement of the large corporations into the research of implementation of Data Minimization methodology (DMM). By involving tech companies into research broader response can be achieved, with more reliable results.	8 - 10 weeks
Literature Review	Review of research papers.	3 -4 weeks
Design phase	Design of DMM used in research with adequate data set, questionnaire, and workshops design. Design of the system development scenario where data governance and DMM can be used.	10 - 12 weeks
Control Study	Workshop with developers tasked with system design without DMM in place to test unconstrained approach to data collection.	1 week
Methodology Review	Review of the control study results, and incorporating findings into methodology design.	2 weeks
Research Study	Workshop with developers tasked with system design using DMM to test its functionality and usability.	1 week
Methodology Review	Review of the research study results, and incorporating findings into methodology design.	2 weeks
Final Study	Workshop with developers tasked with application design using DMM (with assumption that additional changes to methodology improved process of data minimization).	1 week
Results Review	Review and address findings.	4 weeks

Bibliography

Amato F., Coppolino, L., D’Antonio, S., Mazzocca, N., Moscato, F., Sgaglione, L. (2020). An abstract reasoning architecture for privacy policies monitoring. Future Generation Computer Systems 106, 393-400. https://doi.org/10.1016/j.future.2020.01.019.

Antignac T., Le Métayer D. (2014) Privacy Architectures: Reasoning about Data Minimisation and Integrity. In: Mauw S., Jensen C.D. (eds) Security and Trust Management. STM 2014. Lecture Notes in Computer Science, vol 8743, (17-32) . Springer.

Balapour, A., Nikkhah, H., R., Sabherwal, R. (2020). Mobile application security: Role of perceived privacy as the predictor of security perceptions. International Journal of Information Management 52, 102063. https://doi.org/10.1016/j.ijinfomgt.2019.102063.

Barbosa, P., Brito, A., Almeida, H. (2019). Privacy by Evidence: A Methodology to develop privacy-friendly software applications. Information Sciences. Available online 25 September 2019. https://doi.org/10.1016/j.ins.2019.09.040.

Biega, A., J., Potash, P., Daumé III, H., Diaz, F., Finck. M. (2020). Operationalizing the Legal Principle of Data Minimization for Personalization. SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information, 399–408. https://doi.org/10.1145/3397271.3401034.

Chen, X. (2020). A security integration model for private data of intelligent mobile communication based on edge computing. Computer Communications 162, 204-211. https://doi.org/10.1016/j.comcom.2020.08.026.

General Data Protection Regulation (EU), (2016). European Union data protection. https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en. Accessed on: 2020-03-31.

Lee, H. (2020). Home IoT resistance: Extended privacy and vulnerability perspective. Telematics and Informatics 49, 101377. https://doi.org/10.1016/j.tele.2020.101377.

Li, P., Cho, H., Goh, Z., H. (2019). Unpacking the process of privacy management and self-disclosure from the perspectives of regulatory focus and privacy calculus. Telematics and Informatics 41, 114–125. https://doi.org/10.1016/j.tele.2019.04.006.

Bibliography

Mazel, J., Garnier, R., Fukuda, K. (2019). A comparison of web privacy protection techniques. Computer Communications 144, 162-174. https://doi.org/10.1016/j.comcom.2019.04.005.

Mousavi, R., Chen, R., Kim, D., J., Chen, K. (2020). Effectiveness of privacy assurance mechanisms in users' privacy protection on social networking sites from the perspective of protection motivation theory. Decision Support Systems 135, 113323. https://doi.org/10.1016/j.dss.2020.113323.

Pernot-Leplay, E., (2020). China’s Approach on Data Privacy Law: A Third Way Between the U.S. and the EU? Penn State Journal of Law & International Affairs, 8. Available at: https://ssrn.com/abstract=3542820. Accessed on: 13.10.2020.

Sedayao, J. , Bhardwaj, R. , Gorade, N. (2014). Making big data, privacy, and anonymization work together in the enterprise: experiences and issues. 2014 3rd International Congress on Big Data (Big Data Congress), (601–607). https://doi.org/10.1109/BigData.Congress.2014.92.

Senarath, A., Arachchilage, N., A., G. (2018). Understanding user privacy expectations: A software developer’s perspective. Telematics and Informatics 35, 1845-1862. https://doi.org/10.1016/j.tele.2018.05.012.

Senarath, A., Arachchilage, N., A., G. (2019). A data minimization model for embedding privacy into software systems. Computers & Security 87, 101605. https://doi.org/10.1016/j.cose.2019.101605.

Sullivan, C. (2019). EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross border data transfers and protection of personal data in the IoT era. Computer Law & Security Review 35, 380-397. https://doi.org/10.1016/j.clsr.2019.05.004.

Dataset

Dataset for this study is based on GDPR (2016), and all data will be collected during workshops where developers will be presented with system design scenario. Every data field relevance will be described using seven-point Likert scale, ranging from “Strongly Disagree” to “Strongly Agree”.

Personal Data	Type	Description
name	String	Name of the individual
surname	String	Surname of the individual
address	String	Address of the individual
email	String	Email of the individual
mobile_phone	String	Phone number of the individual
age	Int	Age of the individual

Sensitive Personal Data	Type	Description
race	String	Ethnic origin of the individual
religion	String	Religious beliefs of the individual
mental_issues	String	Mental health of the individual
physical_issues	String	Physical health of the individual
sexuality	String	Sexuality of the individual
offence	String	Offence committed or alleged to have been committed

Likert Scale
Strongly Disagree
Disagree
Somewhat Disagree
Neither Agree nor Disagree
Somewhat Agree
Agree
Strongly Agree