POPULAR SCIENTIFIC ABSTRACT

Jens Hjort Schwee

Software Tools for Privacy Control in Publication of Cyber-Physical Data

The development of affordable Internet of Things sensors has made it possible to monitor several aspects of human life. These sensors are among others being used by Cyber-Physical Systems (CPS) as input for the control of systems. One class of such CPS is used for controlling smart buildings. The sensor infrastructure is used to optimize both the energy efficiency and the comfort of the occupants of these buildings. Recent work in the smart building domain is using the collected data to develop and deploy data-driven applications for further optimizations. The owners of the smart building may be interested in sharing some of the collected sensor data with external contractors operating in the building, or as open data, in order for them to train data-driven applications for optimizing their operations. The data sharing process is regulated in a number of privacy laws and regulations, including the European Union's General Data Protection Regulation. To identify the potential privacy implications of sharing the data, an organization should perform a privacy risk analysis identifying the risks for both the monitored occupants and the organization. However, the task of identifying all the related privacy implications is particularly challenging due to advancements enabling many inference and correlation possibilities. Based on the results of the privacy risk analysis the organization must apply appropriate privacy protection on the data before it can be shared. In this Ph. D. dissertation, we explore several aspects of the data sharing process. We have conducted a study identifying problem areas in how State-of-the-Practice methods are used to protect smart building datasets. We found that the methods could not properly protect the explored dataset. Furthermore, we have contributed to the open data pool by publishing a smart building dataset. We have created an ontology, improving the ability to model privacy-related risks in and attacks against datasets. Likewise, we lowered the amount of effort needed to identify privacy-related risks for a specific dataset by designing and evaluating semi-automatic tools. The tools use knowledge from State-of-the-Art methods to identify both the inference and correlation possibilities. It accounts for the type of data, and the spatiotemporal granularity for the identification of the risks. Furthermore, have we designed and evaluated a privacy protection method, which can be applied for data at zone-level granularity with a limited number of each sensor. This Ph. D. dissertation contributes tooling that reduces the amount of information needed to perform privacy risk analyses, enabling more datasets to be properly protected and safely shared.