A Comparative Study on data anonymization techniques to ensure the data privacy

EMSK Ekanayake, D Ganepola

Department of Computer Science, Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka

Abstract. Personal data privacy has become a common topic right now. Because modern technology generates so many public and private datasets, data security has become an unavoidable task. Initially, the priority was given to data security for organizations and businesses, but now it is also necessary to provide security for personal data. The task of processing the data and using it for analysis without compromising the anonymity of the data is critical and must be completed. Data anonymization is the practice of protecting private or secret information by removing or encoding identifiers that link individuals to the stored data. Laws and regulations such as the GDPR were created to unify this anonymity. The concepts of anonymization and data privacy will be explained in this paper. Some of the most common anonymization techniques, as well as an efficient software to support their use, have been discussed. The main goal is to figure out which techniques provide a higher level of anonymity, their strengths and weaknesses, and the benefits of using them. Using existing research papers data has been analysed. Analysing most used data anonymizations algorithms decisions have been made. Suppression, Noise addition, L-Diversity, K-Anonymity, Shuffling have been identified as main anonymization techniques. There is anonymization software that can be used to apply the techniques described systematically. ARX, μ-Argus, SDCMicro, and Privacy Analytics Eclips were discussed. K-Anonymity is identified as a commonly used Anonymization technique. In many real-world applications, more than one technique has been used. As a result, when properly implemented, anonymization techniques ensure data privacy. It may be advantageous to use a combination of techniques in certain circumstances. In many cases, after using anonymization techniques on a dataset, it may be possible to infer information about an individual in some way, even if the inference is not very accurate.

Keywords: *data privacy, Anonymization's techniques, data anonymization tools*