Defining the Process of Pseudonymity for PROV

Jenni Reuben[1], Heather Packer[2], Hans Hedbom[1], and Simone Fischer-Hübner[1]

[1] Karlstad University, Sweden
[firstname.lastname]@kau.se
[2] University of Southampton, UK
{hp3, l.moreau}@ecs.soton.ac.uk

Introduction

Provenance is a well known concept in the art world, it refers to the documented history of an art object, which is used to evaluate the significance of the art object in relation to other similar objects [3]. Similarly data provenance[3] - a record which describes the derivation history of a digital artifact asserts the trustworthiness of the digital artifact. Provenance describes how people, institutions, processes and other data items influence or involve in the generation, modification and delivery of a digital artifact [1].

Conceptually, provenance of a data could trace back to the origin of the Universe [2]. In today's computing, information is processed across multiple distributed systems using various different technologies. Hence in order to support a meaningful provenance analysis concerning a piece of data or thing, an interoperable exchange of provenance information across systems is desirable. PROV-DM - a W3C recommendation [1] allows heterogeneous systems to express and exchange its provenance information using a generic data model.

Nevertheless provenance information which describes how a data object arrived to its current state would consequently constitute personal identifiers, quasi identifiers and sensitive information. Preliminary results of our earlier work[4] lists out various privacy threats that arise in provenance-aware applications. In this work we present novel solutions to some of the identified privacy threats. In particular, personal data exposure threats in provenance information. We propose the use of pseudonyms to veil the exposed personal information in provenance. The research challenge is to guarantee stronger anonymity while still be able to link provenance information for provenance analysis. In this paper we identify the requirements for a pseudonymization solution in provenance context, propose a model for pseudonymization solution and evaluation of our model using stronger attacker model(s).

References

1. Missier, P., Moreau, L.: PROV-dm: The PROV data model. W3C recommendation, W3C (Apr 2013), http://www.w3.org/TR/2013/REC-prov-dm-20130430/

[3] Hereafter, will simply referred to as provenance

[4] Under submission - ARES ISPM 2016

2. Moreau, L.: The Foundations for Provenance on the Web. Found. Trends Web Sci. 2, 99–241 (Feb 2010)
3. Munroe, S., Miles, S., Moreau, L., Vázquez-Salceda, J.: PrIMe: A Software Engineering Methodology for Developing Provenance-aware Applications. In: Proceedings of the 6th International Workshop on Software Engineering and Middleware. pp. 39–46. SEM '06, ACM, New York, NY, USA (2006), http://doi.acm.org/10.1145/1210525.1210535