Accepted Manuscript

Publishing privacy logs to facilitate transparency and accountability

Reza Samavi, Mariano P. Consens

PII: S1570-8268(18)30008-8
DOI: https://doi.org/10.1016/j.websem.2018.02.001
Reference: WEBSEM 455

To appear in: *Web Semantics: Science, Services and Agents on the World Wide Web*

Received date : 22 December 2016
Revised date : 23 November 2017
Accepted date : 15 February 2018

Please cite this article as: R. Samavi, M.P. Consens, Publishing privacy logs to facilitate transparency and accountability, *Web Semantics: Science, Services and Agents on the World Wide Web* (2018), https://doi.org/10.1016/j.websem.2018.02.001

Publishing Privacy Logs to Facilitate Transparency and Accountability

Reza Samavi[a,*], Mariano P. Consens[b]

[a] *Department of Computing and Software, McMaster University, 1280 Main St. West, Hamilton, Ontario L8S 4K1, Canada*
[b] *Information Engineering, MIE, University of Toronto, 5 King's College Road, Toronto, Ontario M5S 3G8, Canada*

Abstract

Compliance with privacy policies imposes requirements on organizations and their information systems. Maintaining auditable privacy logs is one of the key mechanisms employed to ensure compliance, but the logs and their auditing reports are designed and implemented on an application by application basis. This paper develops a Linked Data model and ontologies to facilitate the sharing of logs that support privacy auditing and information accountability among multiple applications and participants. The L2TAP modular ontologies accommodate a variety of privacy scenarios and policies. SCIP is the key module that synthesizes contextual integrity concepts and enables query based solutions that facilitate privacy auditing. Other L2TAP modules describe logs, participants, and log events, all identified by web accessible URIs and include relevant provenance information to support accountability. A health self-management scenario is used to illustrate how privacy preferences, accountability obligations, and access to personal information can be published and accessed as linked data by multiple participants, including the internal and external auditors. We contribute query based algorithmic solutions for two fundamental privacy auditing processes that analyze L2TAP logs: obligation derivation and compliance checking. The query based solutions that we develop require SPARQL implementations with limited RDFS reasoning power, and are therefore widely supported by commercial and open source systems. We also provide experimental validation of the scalability of our query based solution for compliance checking over L2TAP logs.

Keywords: Privacy, Policy, Audit Log, Accountability, Linked Data, Semantic Web, Ontology

1. Introduction

The protection of individuals' privacy is becoming increasingly more challenging in the era of social computing and data driven science. An important aspect of privacy protection is *information accountability*, ensuring the policies that govern the lifecycle of personal information (i.e., collect, use, transform, and share users' data) are respected by all parties who are involved in the process [1, 2]. Ensuring compliance is a complex task involving multiple participants including the data subjects whose data are at stake and have privacy preferences, data collectors who will be liable if the privacy policies are violated, and internal and external auditors who oversee collection and usage of personal information. Privacy auditing supports information accountability by *validation* (verifies a posteriori if a participant has performed the tasks as expected), *attribution* (finds the responsible participant in case of a deviation from policies), and *evidence* (produces evidence that can be used to convince an auditor if a fault has or has not occurred)[3, 1]. This paper presents a Linked Data [4] oriented model that facilitates privacy auditing.

The need for privacy auditing is even more important in the era of the personal web [5] where users are empowered to mix and match a variety of web resources and services to achieve their personal goals [6]. For example, consider the *Sharing Data with Fitness Coach* use case described in [7] (Fig. 1). Mary is interested in self-managing her blood pressure using Personal Health Record (PHR) services. PHR systems (e.g., Microsoft HealthVault [8]) are open platforms with published application programming interfaces (API) that allow users to utilize an extensible ecosystem of personal health applications [9]. Mary adds a blood pressure collec-

*Corresponding author: Tel: +1 905 525-9140 x. 24895
Email address: samavir@mcmaster.ca (Reza Samavi)