International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

PRIVACY AWARE PERSONAL DATA STORAGE PPDS LEARNING HOW TO PROTECT USER PRIVACY FROM APPLICATIONS

G. Chakrapani[1], C. Durga Dhanush Naidu[2], Sathwika Reddy[2], CH. Priyanka[2], B. Neethi Kireeta[2], Bhupender Singh Raj[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of Computer Science Engineering

[1,2]Malla Reddy Engineering College and Management Sciences, Medchal, Hyderabad

ABSTRACT

This paper aims to design a Privacy-aware Personal Data Storage (P-PDS) that automatically takes privacy-aware decisions on third-party access requests in accordance with user preferences. PDS has moved from a service-centric to a user-centric model, enabling individuals to store and control their personal data in a unique logical repository. The proposed P-PDS is based on preliminary results, where it has been demonstrated that semi-supervised learning can be successfully exploited to make a PDS able to automatically decide whether an access request has to be authorized or not.The key issue of helping users specify their privacy preferences on PDS data has not been deeply investigated, as average PDS users are not skilled enough to understand how to translate their privacy requirements into a set of privacy preferences. Studies have shown that average users might have difficulties in properly setting potentially complex privacy preferences.To help users protect their PDS data, the authors evaluated the use of different semi-supervised machine learning approaches for learning privacy preferences of PDS owners. They found that ensemble learning was the best fit for the considered scenario. However, the design of a Privacy-aware Personal Data Storage requires further investigation, as it still requires many interactions with PDS owners to collect a good training dataset.

Keywords: Data Storage, User Privacy, PPDS Learning.

1. INTRODUCTION

Nowadays personal data we are digitally producing are scattered in different online systems managed by different providers (e.g., online social media, hospitals, banks, airlines, etc). In this way, on the one hand users are losing control on their data, whose protection is under the responsibility of the data provider, and, on the other, they cannot fully exploit their data, since each provider keeps a separate view of them. To overcome this scenario, Personal Data Storage (PDS) [2]–[4] has inaugurated a substantial change to the way people can store and control their personal data, by moving from a service-centric to a user-centric model. PDSs enable individuals to collect into a single logical vault personal information they are producing. Such data can then be connected and exploited by proper analytical tools, as well as shared with third parties under the control of end users. This view is also enabled by recent developments in privacy legislation and, in particular, by the new EU General Data Protection Regulation (GDPR), whose art. 20 states the right to data portability, according to which the data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format, thus making possible data collection into a PDS.Up to now, most of the research on PDS has focused on how to enforce user privacy preferences and how to secure data when stored into the PDS (see Section 7 for more details). In contrast, the key issue of helping users to specify their privacy preferences on PDS data has not been so far deeply investigated. This is a fundamental issue since average PDS users are not skilled enough to understand how to translate their privacy requirements into a set of privacy preferences. As several studies have shown, average users might have difficulties in properly setting potentially complex privacy preferences [5]–[7]

2. EXISTING SYSTEM

Oort [27] is a user-centric cloud storage system that organizes data by users rather than applications, considering global queries which find and combine relevant data fields from relevant users. Moreover, it allows users to choose which applications can access their own data, and which types of data to be shared with which users. Sieve [28] allows user to upload encrypted data to a single cloud storage. It utilizes key- homomorphic scheme to provide cryptographically enforced access control.

Amber [29] has proposed an architecture where users can choose applications to manipulate their data but it does not mention either how the global queries work or how the application providers interact with. In [2], authors developed a user-centric framework that share with third parity only the answers to a query instead of the raw data. Mortier et al. [30] have proposed a trusted platform called Databox, which can manage personal data by a fine grained access control mechanism but do not focus on policy learning. Recently, [31] proposed a Block chain-based Personal Data Store (BC-PDS) framework, which leverages on BlockChain to secure the storage of personal data. However, all the above proposals focus on access control enforcement, whereas they do not consider user preference or policy learning.

Privacy preference enforcement have been also investigated in different domains, such as for instance social networks where most of the platforms offer users a privacy setting page to manually set their privacy preferences. Research works have tried to alleviate theburden of this setting, by exploiting machine learning tools. For instance, [32], [33] have investigated the use of semi-supervised and unsupervised approaches to automatically extract privacy settings in social media. In [34], authors have considered location based data. They have compared the accuracy of manually set privacy preferences with the one of an automated mechanism based on machine learning. The results show that machine learning approaches provide better result than user-defined policies. Bilogrevic et al. [35] also present a privacy preference framework that (semi)automatically predicts sharing decision, based on personal and contextual features. The authors focus only on g location information.

3. PROPOSED SYSTEM

The system proposes a revised version of the ensemble learning algorithm proposed in [1], to enforce a more conservative approach w.r.t. users privacy. In particular, we reconsider how ensemble learning handles decisions for access requests for which classifiers return conflicting classes. In general, the final decision is taken selecting the class with the highest aggregated probabilities. However, this presents the limit of not considering user perspective, in that, it does not take into account which classifier is more relevant for the considered user.

To cope with this issue, we propose an alternative strategy for aggregating the class labels returned by the classifiers. According to this approach, we assign a personalized weight to each single classifier used in ensemble learning. We also show how it is possible to learn these weights from the training dataset, without the need of further input from the P-PDS owner. Experiments show that this approach increases users satisfaction as well as the learning effectiveness.

4. CONCLUSION

This paper proposes a Privacy-aware Personal Data Storage, able to automatically take privacy-aware decisions on third parties access requests in accordance with user preferences. The system relies on active learning complemented with strategies to strengthen user privacy protection. As discussed in the paper, we run several experiments on a realistic dataset exploiting a group of 360 evaluators. The obtained results show the effectiveness of the proposed approach. We plan to extend this work along several directions. First, we are interested to investigate how P-PDS could scale in

the IoT scenario, where access requests decision might depend also on contexts, not only on user preferences. Also, we would like to integrate P-PDS with cloud computing services (e.g., storage and computing) so as to design a more powerful P-PDS by, at the same time, protecting users privacy.

REFERENCES

[1]. B. C. Singh, B. Carminati, and E. Ferrari, "Learning privacy habits of pds owners," in Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on. IEEE, 2017, pp. 151–161.

[2]. Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, "openpds: Protecting the privacy of metadata through safeanswers," PloS one, vol. 9, no. 7, p. e98790, 2014.

[3]. B. M. Sweatt et al., "A privacy-preserving personal sensor data ecosystem," Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

[4]. B. C. Singh, B. Carminati, and E. Ferrari, "A risk-benefit driven architecture for personal data release," in Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on. IEEE, 2016, pp. 40–49.

[5]. M. Madejski, M. Johnson, and S. M. Bellovin, "A study of privacy settings errors in an online social network," in Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on. IEEE, 2012, pp. 340–345.

[6]. L. N. Zlatolas, T. Welzer, M. Hericko, and M. H ˇ olbl, "Privacy an- ¨ tecedents for sns self-disclosure: The case of facebook," Computers in Human Behavior, vol. 45, pp. 158–167, 2015.

[7]. D. A. Albertini, B. Carminati, and E. Ferrari, "Privacy settings recommender for online social network," in Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on. IEEE, 2016, pp. 514–521.

[8]. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in International workshop on privacy enhancing technologies. Springer, 2006, pp. 36– 58.

[9]. R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in Proceedings of the 2005 ACM workshop on Privacy in the electronic society. ACM, 2005, pp. 71–80.

[10]. Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Analyzing facebook privacy settings: user expectations vs. reality," in Proceedings of the 2011 ACM SIGCOMM

[11]. P. Nyoni and M. Velempini, "Privacy and user awareness on facebook," South African Journal of Science, vol. 114, no. 5-6, pp. 27–31, 2018.

[12]. R. Polikar, "Ensemble learning," in Ensemble machine learning. Springer, 2012, pp. 1–34.

[13]. T. G. Dietterich et al., "Ensemble methods in machine learning," Multiple classifier systems, vol. 1857, pp. 1–15, 2000.

[14]. B. Settles, "Active learning literature survey," University of Wisconsin, Madison, vol. 52, no. 55-66, p. 11, 2010.

[15]. X. Zhu, "Semi-supervised learning literature survey," Computer Science, University of Wisconsin-Madison, vol. 2, no. 3, p. 4, 2006