Privacy-Aware Personal Data Storage (P-PDS) Learning How To Protect User Privacy From External Applications

Borra Sravani , Sri.V.Bhaskara Murthy
MCA Student, Associate Professor
Dept Of MCA
B.V.Raju College, Bhimavaram

ABSTRACT

Recently, Personal Data Storage (PDS) 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. PDS offers individuals the capability to keep their data in a unique logical repository, that can be connected and exploited by proper analytical tools, or shared with third parties under the control of end users. 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. In contrast, in this paper we aim at designing a Privacy-aware Personal Data Storage (P-PDS), that is, a PDS able to automatically take privacy-aware decisions on third parties access requests in accordance with user preferences. The proposed P-PDS is based on preliminary results presented in [1], 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. In this paper, we have deeply revised the learning process so as to have a more usable P PDS, in terms of reduced effort for the training phase, as well as a more conservative approach w.r.t. users privacy, when handling conflicting access requests. We run several experiments on a realistic dataset exploiting a group of 360 evaluators. The obtained results show the effectiveness of the proposed approach.

I. 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]. For example, let us consider Facebooks privacy setting, where users need to configure the options manually according to their desire. In [8], [9], authors survey users awareness, attitudes and privacy concerns on profile information and find that only a small number of users change the default privacy preferences on Facebook. Interestingly, in [10], authors find that even when users have changed their default privacy settings, the modified settings do not match the expectations (these are reached only for 39% of users). Moreover, another survey in [11] has shown that Facebook users are not aware enough on protection tools that designed to protect their personal data. According to their study the majority (about 88%) of users had never read the Facebook privacy policy. To help users on protecting their PDS data, in [1], we have evaluated the use of different semi-supervised machine learning approaches for learning privacy preferences of PDS owners. The idea is to find a learning algorithm that, after a training period by the PDS owner, returns a classifier able to automatically decide if access requests submitted by third parties are to be authorized or denied. In [1], we have shown that, among different semi-supervised learning approaches, the one that better fits the considered scenario is ensemble learning [12], [13] (see Section 2 for more details). Even though the identification of the learning approach is an essential step, the design of a Privacyaware Personal Data Storage (P-PDS), that is, a PDS able to automatically take privacy-aware decisions on third parties access requests requires further investigation. One critical aspect to consider is the usability of the system. Even if semi-supervised techniques require less users effort, compared to manually setting privacy preferences, they still require many interactions with PDS owners to collect a good training dataset.

II. 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 the burden 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.

Disadvantages

- In the existing work, the system doesn't have strong techniques to implement Privacy-aware Personal Data Storage (P-PDS).
- The system doesn't have active learning which is to select from the training dataset the most representative instances to be labeled by users.

III. PROPOSED SYSTEM

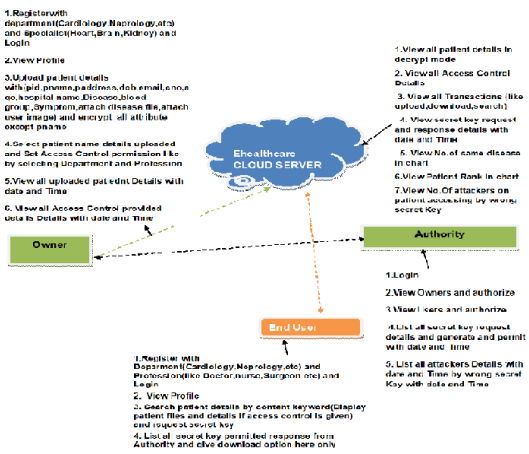
The system proposes a revised version of the ensemble learning algorithm proposed in [1], to enforce a more conservative approach w.r.t. users privacy. In particular, we reconsider how ensemble learning handles decisions for access requests for which classifiers return conflicting classes. In general, 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, thus 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.

Advantages

- PDS able to automatically take privacy-aware decisions on third parties access requests requires further investigation.
- The system proposes a revised version of the ensemble learning algorithm proposed in this system, to enforce a more conservative approach w.r.t. users privacy.

IV. SYSTEM ARCHITECTURE



V. IMPLEMENTATION

DATA OWNER

In this module, Data owner has to register to cloud and logs in, Encrypts and uploads a file to cloud server and also performs the following operations such as Register with department(Cardiology,Neprology,etc) and Specialist(Heart,Brain,Kidney) and Login and View Profile ,Upload patient details

with(pid,pname,paddress,dob,email,cno,age,hospital name,Disease,blood group,Symptom,attach disease file, attach user image) and encrypt all attribute except pname ,Select patient name details uploaded and Set Access Control permission like by selecting Department and Profession and View all uploaded patient Details with date and Time ,View all Access Control provided details with date and Time.

Ehealthcare

CLOUD SERVER

In this module the cloud will authorize both the owner and the user and also performs the following operations such as View all patient details in decrypt mode and View all Access Control Details, View all Transactions (like upload, download, search) and View secret key request and response details with date and Time View No.of same disease in chart, View Patient Rank in chart and View No.Of attackers on patient accessing by wrong secret Key

Authority

In this module, the Authority performs the following operations such as Login ,view Owners and authorize and View Users and authorize,List all secret key request details and generate and permit with date and Time and List all attackers Details with date and Time by wrong secret Key with date and Time.

End USER

In this module, the user has to register to cloud and log in and performs the following operations such as Register with Deparment(Cardiology,Neprology,etc) and Profession(like Doctor,nurse,Surgeon etc) and Login ,View Profile and Search patient details by content keyword(Display patient files and details if access control is given) and request secret key and List all secret key permitted response from Authority and give download option here only.

VI. CONCLUSION

This paper proposes a Privacy-aware Personal Data Storage, able to automatically take privacyaware 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 selfdisclosure: 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] A. 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 conference on Internet measurement conference. ACM, 2011, pp. 61–70.

[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.