ENHANCING USER DATA STORAGE AND PROTECTION USING PRIVACY PROTECTION TECHNIQUES

1Mr. R. MADHAVA REDDY, 2P. VARSHA, 3J. AKSHITH, 4V. HARSHAVARDHAN

1Assistant Professor, Department of AI&DS, Sri Indu College of Engineering and Technology-Hyderabad

234Under Graduate, Department of AI&DS, Sri Indu College of Engineering and Technology-Hyderabad

ABSTRACT

Recently, Personal Data Storage (PDS) has initiated 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 ability to keep their data in a distinct logical repository, that can be connected and utilize 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 put up on user privacy preferences and how to secure data when stored into the PDS. In contrast, in this project 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 conforming to user preferences. In this project, 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 with respect to 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.

Keywords: Data Storage and Protection, Privacy Protection Techniques

INTRODUCTION

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, 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. For example, let us consider Facebooks privacy setting, where users need to configure the options manually according to their desire. In 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

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 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, we have shown that, among different semi-supervised learning approaches, the one that better first the considered scenario is ensemble learning. Even though the identification of the learning approach is an essential step, the design of a Privacy-aware 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.

To further reduce the required user effort, in the current paper, we leverage on active learning (AL) to minimize user burden for getting the training dataset by, at the same time, achieving better accuracy in determining user privacy preferences. The main idea of active learning is to select from the training dataset the most representative instances to be labeled by users. Literature offers several methods driving the selection of these new instances. The most commonly adopted method is uncertainty sampling. According to this approach, to be labeled by human annotators, active learning selects those instances for which it is highly uncertain how to label them according to the preliminary built model. As this improvement brings benefits in term of accuracy and usability. Additionally, to further improve the performance of the system, we define an alternative uncertainty sampling strategy, which is based on the observation that, for taking a privacy-related decision, some fields of access requests (i.e., data consumer and type of service requesting the data) are more informative than others. Thus, if a new access request presents new values for these fields, the system pushes for a new training (i.e., asking data owner a label for the access request). To enforce this behavior, we introduce a penalization of the uncertainty measure based on the distance of the new access request w.r.t. the access requests previously labeled by the P-PDS owner (we call this strategy history-based active learning). As it will show in the experiments, history-based active learning shows better results than AL in terms of users satisfaction. As a further improvement, in this paper, we propose a revised version of the ensemble learning

algorithm proposed in, 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.

Certainly! This text discusses the concept of Personal Data Storage (PDS) and its significance in the modern digital landscape. It highlights how personal data is currently scattered across various online platforms managed by different providers, leading to a loss of control for users and hindering the full exploitation of their data. PDS aims to address this issue by shifting from a service-centric to a user-centric model, allowing individuals to aggregate their personal information in a centralized vault. This aggregated data can then be analyzed and shared under the user's control.

LITERATURE SURVEY

1. Title: Learning privacy habits of pds owners.

Author: B. C. Singh, B. Carminati, and E. Ferrari.

Abstract: The concept of Personal Data Storage (PDS) has recently emerged as an alternative and innovative way of managing personal data w.r.t. the service-centric one commonly used today. The PDS offers a unique logical repository, allowing individuals to collect, store, and give access to their data to third parties. The research on PDS has so far mainly focused on the enforcement mechanisms, that is, on how user privacy preferences can be enforced. In contrast, the fundamental issue of preference specification has been so far not deeply investigated. In this paper, we do a step in this direction by proposing different learning algorithms that allow a fine- grained learning of the privacy aptitudes of PDS owners. The learned models are then used to answer third party access requests. The extensive experiments we have performed show the effectiveness of the proposed approach.

2. Title: open pds : Protecting the privacy of meta data through safe answers.

Author: Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland.

Abstract: The rise of smartphones and web services made possible the large-scale collection of personal metadata. Information about individuals' location, phone call logs, or web-searches, is collected and used intensively by organizations and big data researchers. Metadata has however yet to realize its full potential. Privacy and legal concerns, as well as the lack of technical solutions for personal metadata management is preventing metadata from being shared and reconciled under the control of the individual. This lack of access and control is furthermore fueling growing concerns, as it prevents individuals from understanding and managing the risks associated with the collection and use of their data. Our contribution is two-fold: we describe openPDS, a personal metadata management framework that allows individuals to collect, store, and give fine-grained access to their metadata to third parties. It has been implemented in two field studies we introduce and analyze SafeAnswers, a new and practical way of protecting the privacy of metadata at an individual level. SafeAnswers turns a hard anonymization problem into a more tractable security one. It allows services to ask questions whose answers are calculated against the metadata instead of trying to anonymize individuals' metadata. The dimensionality of the data shared with the services is reduced from high- dimensional metadata to low-dimensional answers that are less likely to be re-identifiable and to contain sensitive information. These answers can then be directly shared individually or in aggregate. openPDS and SafeAnswers provide a new way of dynamically protecting personal metadata, thereby supporting the creation of smart data-driven services and data science research.

3. Title: A privacy-preserving personal sensor data ecosystem.

Author: B. M. Sweatt et al.

Abstract: Despite the ubiquity of passively-collected sensor data (primarily attained via smartphones), there does not currently exist a comprehensive system for authorizing the collection of such data, collecting, storing, analyzing, and visualizing it in a manner that preserves the privacy of the user generating the data. This thesis shows the design and implementation of such a system, named openPDS, from both the client and server perspectives. Two server-side components are implemented: a centralized registry server for authentication and authorization of all entities in the system, and a distributed Personal Data Store that allows analysis to be run against the stored sensor data and aggregated across multiple Personal Data Stores in a privacy-preserving fashion. The client, implemented for the Android mobile phone operating system, makes use of the Funf Open Sensing framework to collect data and adds the ability for users to authenticate against the registry server, authorize third-party applications to analyze data once it reaches their Personal Data Store, and finally, visualize the result of such analysis within a mobile phone or web browser. A number of example

quantified-self and social applications are built on top of this framework to demonstrate feasibility of the system from both development and user perspectives.

4. Title: A risk-benefit driven architecture for personal data release.

Author: B. C. Singh, B. Carminati, and E. Ferrari.

Abstract: Personal data storages (PDSs) give individuals the ability to store their personal data in a data unified repository and control release of their data to data consumers. Being able to gather personal data from different data sources (e.g., banks, hospitals), PDSs will play strategic role in individual privacy management. As such, PDS demands for new privacy models for protecting personal data. In this paper, we propose a new technical approach that empowers individuals to better control data in PDS. Particularly, we present a privacy-aware PDS architecture by focusing on two logical data zones based on the categories of personal data. Moreover, we propose a strategy for regulating personal data release that takes in consideration both user preferences and possible risks and benefits of the data release.

SYSTEM ANALYSIS

EXISTING SYSTEM

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) 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.

DISADVANTAGES

- Concerns About Data Security although the Personal Data Storage (PDS) concept seeks to provide individuals more data control, security issues could arise.personal information in one place increases the danger that it may be stolen, which might lead to data breaching

- Fragmented Data Ownership Users can collect and manage their own data with the use of PDS, which results in fragmented data ownership This creates challenges in data accuracy and impact on data inconsistency and difficult to maintain accurate and up-to- date information.

PROPOSED SYSTEM

The proposal discussed in demonstrates that semisupervised ensemble learning can be exploited to train a classifier so as to make a PDS able to automatically decide whether an access request has to be authorized or not. However, to build a classifier using a predictive learning model, it is essential to label an initial set of instances, called the training dataset. It is matter of fact that obtaining a sufficient number of labeled instances is time consuming and costly due to the required human input. On the other hand, the size and quality of the training dataset impact the accuracy the classifier might reach. Therefore, Active learning (AL) may be exploited to reduce the size of the training dataset. The key idea of AL is to build the training dataset by properly selecting a reduced number of instances from unlabeled items, rather than randomly choosing them as done by traditional supervised learning algorithms. This makes it possible to efficiently exploit unlabeled instances for developing effective prediction models as well as to reduce the time and cost of labeling.

ADVANTAGES

- Efficient Training Data Selection Active learning (AL) offers the advantage of intelligently selecting training data AL picks unlabeled instances to provide most informative value This approach ensures learner more effectively achieving higher accuracy Reduced Time and Cost of Labeling AL addresses the challenge of acquiring labeled data, which is often time- consuming and costly AL optimizes human labeling resources This reduces overall time and

effort needed for labeling, making it a cost-effective it is an alternative for traditional supervised learning methods.

IMPLEMENTATION AND RESULTS

MODULE DESCRIPTION OWNER

In this application the owner is one of the main module for uploading the files and view the uploads file which are uploaded by the owner before do all these operations the owner should register with the application and the owner should authorized by the cloud.

The Owner module is a core component of the application. It allows the owner to upload files and view the files they've uploaded. However, before performing these actions, the owner needs to register with the application and gain authorization from the cloud. This ensures that only authorized owners can use the system.

USER

In this application the user also a modules to perform the bloom filter operation to access the files from the cloud, before do the search operations the user should get the search permission from the cloud then only the user can search the files after get the details of the searched file, if the user want to download the user should get the trapdoor key from the trapdoor Generator, then the user can able to download the file. To do all these operation the user should register with application and the user should accessed by the cloud .The User module is responsible for performing bloom filter operations to access files stored in the cloud. Before conducting any search operations, the user must obtain search permission from the cloud. Once granted, the user can search for files. If they wish to download a file, they need to acquire a trapdoor key from the Trapdoor Generator. This key enables them to download the file securely. Like the owner, users need to register with the application and gain access approval from the cloud.

TRAPDOOR GENERATOR

The trapdoor is used to generate the trapdoor key for the requested users. Here the trapdoor should login directly with the application.

The Trapdoor Generator is a crucial component responsible for generating trapdoor keys. These keys are essential for secure file access. The generator has a direct login to the application, allowing it to perform its key generation function efficiently.

CLOUD

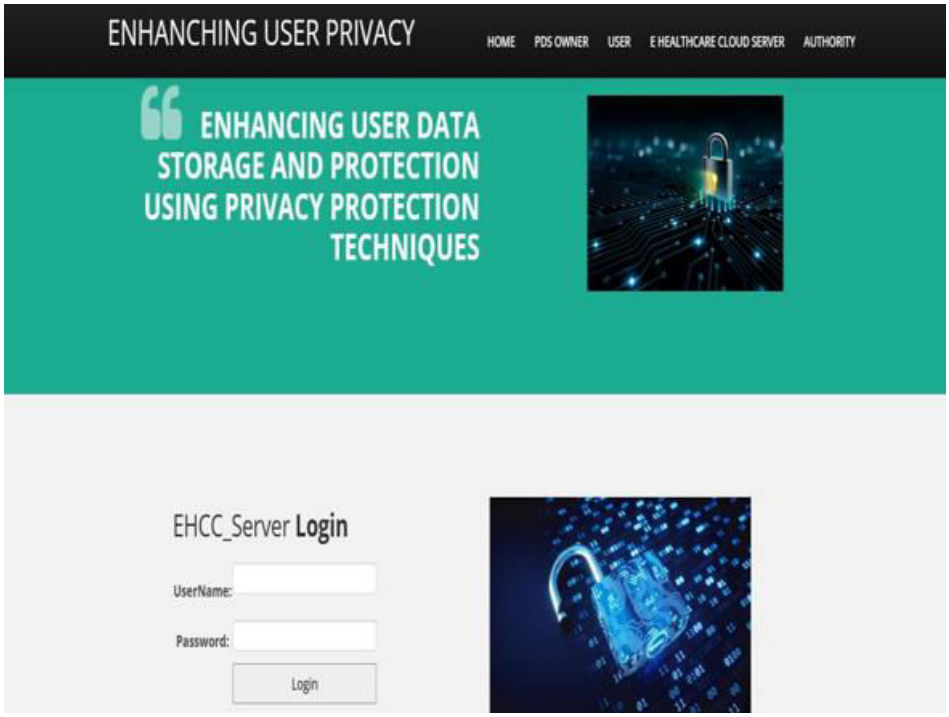
The cloud is the main module to operate this project in the users activation owner activation and also the cloud can check the following operations like search permission provides to the users, can check

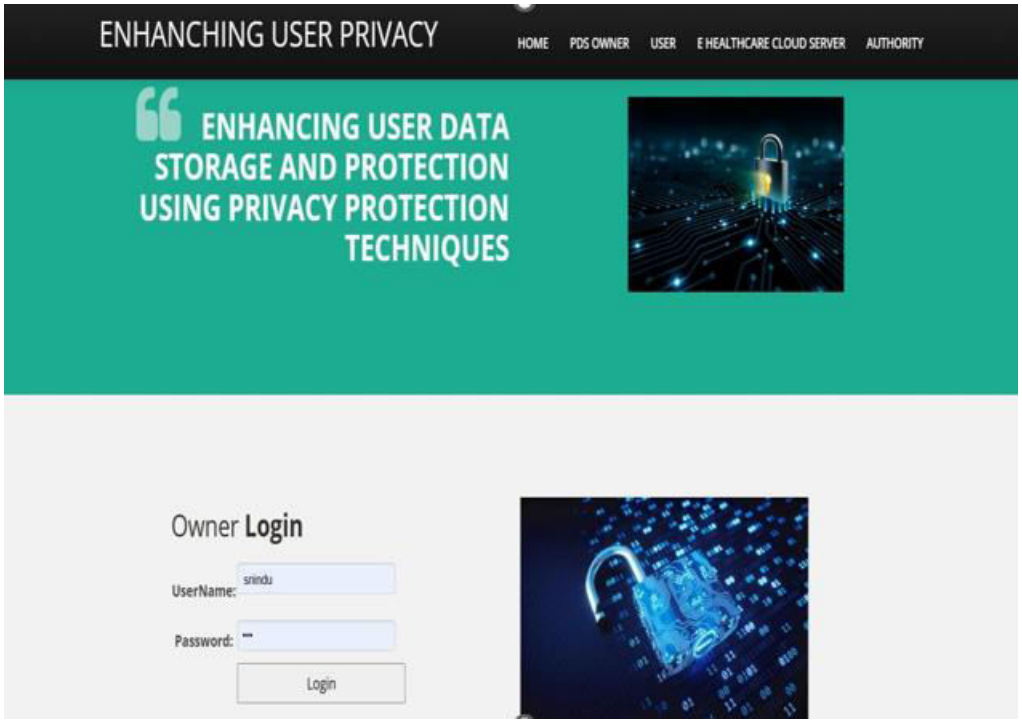
the top-k searched keyword, top-k similarity in chart, top-k searched keyword in chart. Primarily the cloud should login. Then only the cloud can perform the above mentioned actions. The Cloud module serves as the central platform for managing the project. It oversees user and owner activations. The cloud also handles various operations, including granting search permissions to users, monitoring top-k searched keywords, providing similarity data in charts, and tracking popular keywords. To perform these tasks, the cloud must log in securely to the application.

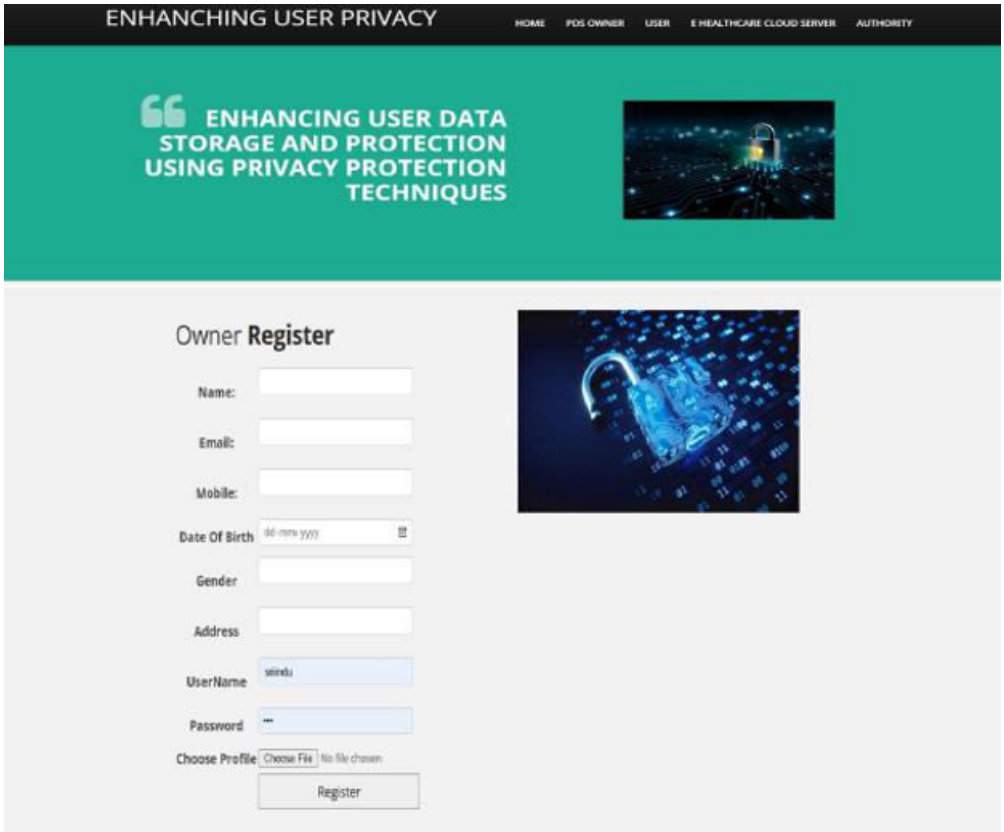
ATTACKER

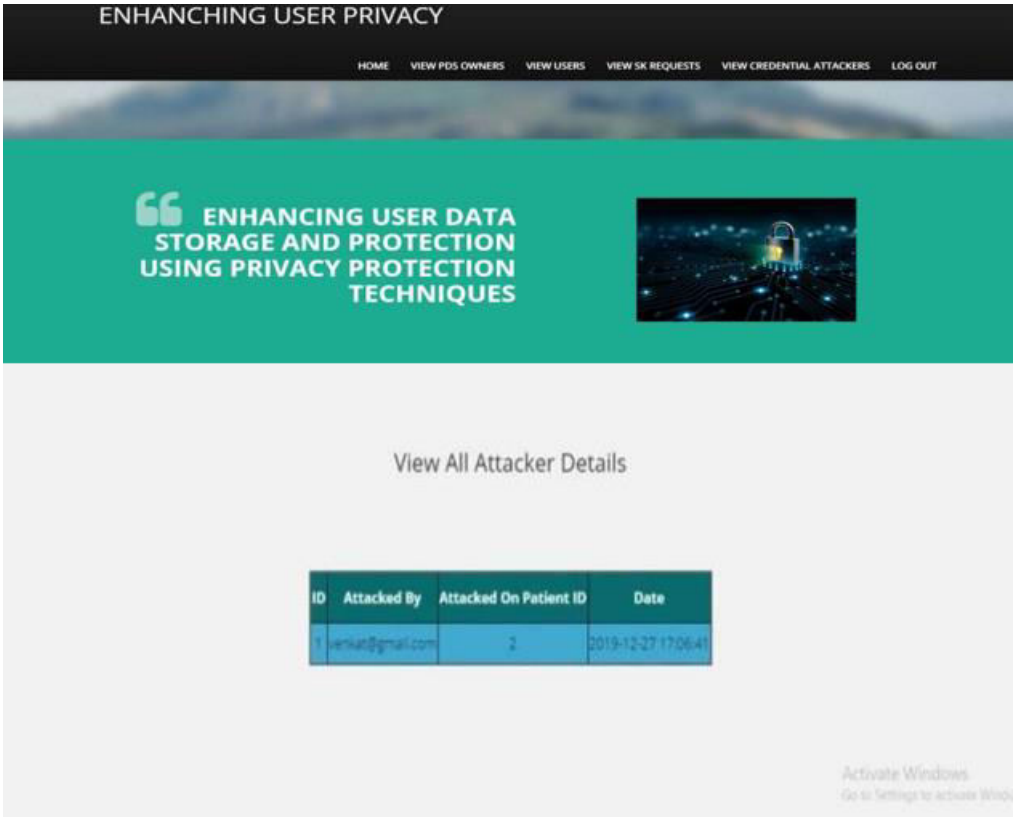
The attacker is the unauthorized perform to attack the owner files.The Attacker module represents an unauthorized entity attempting to breach the security of the owner's files. This individual poses a threat to the integrity and confidentiality of the uploaded files. It's important for the system to have robust security measures in place to thwart potential attacks and protect sensitive data.

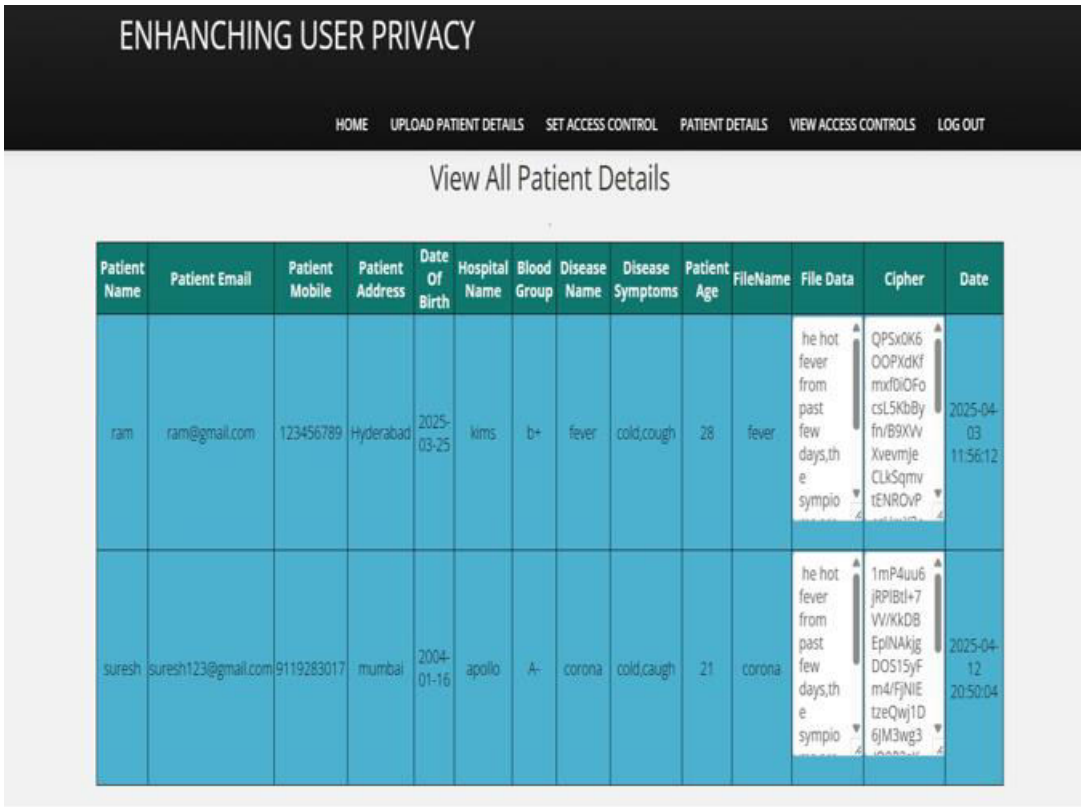
These modules collectively form a comprehensive system for secure file management and access control. Each module plays a distinct role, ensuring that only authorized users can interact with the application and its resources. Additionally, the system includes measures to defend against potential threats from unauthorized entities.




ENHANCING USER PRIVACY
HOME
PDS OWNER
USER
E HEALTHCARE CLOUD SERVER
AUTHORITY
ENHANCING USER DATA STORAGE AND PROTECTION USING PRIVACY PROTECTION TECHNIQUES
Owner Login
UserName:
sindu
Password:
Login


ENHANCING USER PRIVACY
HOME
PDS OWNER
USER
E HEALTHCARE CLOUD SERVER
AUTHORITY
ENHANCING USER DATA STORAGE AND PROTECTION USING PRIVACY PROTECTION TECHNIQUES
Owner Register
Name:
Email:
Mobile:
Date Of Birth
Gender
Address
UserName
sindu
Password
Choose Profile
Choose File
No file chosen
Register





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.

In conclusion, this project presents a robust IoT security solution integrating advanced cryptography and steganography techniques. By prioritizing data privacy and user authentication, it addresses critical vulnerabilities in IoT networks. The proposed system, with its Adaptive Firefly optimization algorithm, showcases promising results in securing confidential medical data. Looking forward, the project holds potential for further advancements in security measures, making significant contributions to the evolving landscape of IoT technology. Its adaptability and scalability offer a solid foundation for future research and applications across diverse industries.

FUTURE SCOPE

The proposed IoT security solution demonstrates a robust potential for future advancements across various domains. Firstly, it holds the capability to adapt to evolving cyber threats by integrating state-of-the-art cryptographic techniques and steganographic methods, ensuring a proactive defense against potential attacks. Moreover, exploring integration with emerging technologies like blockchain can provide decentralized authentication and data integrity verification, further fortifying the security framework. Implementing machine learning and artificial intelligence algorithms can bolster anomaly detection, enabling the system to dynamically adapt and respond to evolving threats intelligently. Adapting the solution for edge and fog computing environments signifies a strategic move towards optimizing resource utilization and reducing latency, particularly crucial for real-time IoT applications.

Expanding the project to encompass comprehensive device management and security orchestration capabilities offers a holistic approach to IoT security, ensuring a cohesive and well- coordinated defense. Extending the security framework to diverse industries such as smart homes, industrial automation, and smart cities amplifies its impact, demonstrating its versatile applicability. Establishing industry standards and seeking certification instills trust and encourages widespread adoption of the proposed security solution. Furthermore, integrating incident response and forensics capabilities equips the system to swiftly identify and mitigate security incidents, bolstering overall resilience.

Adapting to evolving data protection laws and compliance requirements ensures that the system remains in alignment with legal frameworks, enhancing its credibility and trustworthiness. Encouraging collaborative research efforts and contributing to open-source projects fosters innovation, driving continuous improvement and community-driven enhancements. Finally, the development of educational resources and awareness programs empowers users to take proactive measures in securing their IoT devices and data, further enhancing the overall efficacy of the security solution. Collectively, these potential avenues highlight a promising future for the proposed IoT security solution, encompassing a wide array of enhancements across technology, compliance, and user education domains.

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, 2017.

[3] B. M. Sweatt et al., “A privacy-preserving personal sensor data ecosystem,” Ph.D. dissertation, Massachusetts Institute of Technology, 2017.

[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, 2015, pp. 340–345.

[6] L. N. Zlatolas, T. Welzer, M. Heri˘cko, and M. H¨olbl, “Privacy antecedents 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, 2014, pp. 514–521.

[8] A. Acquisti and R. Gross, “Imagined communities: Awareness, information sharing, and privacy on the face book,” 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, 2005, pp. 61–70.