|| Volume 5 || Issue 11 || November 2020 || ISSN (Online) 2456-0774

INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

SAFEGUARDING PRIVACY AND ACCOMPLISHING DATA TRUTHFULNESS IN DATA MARKETS

Mohd Meraj Ali [1] , Abdul Rasool MD [2]

Research Scholar, Dept. of Computer Science & Engineering, LIET, Hyderabad[1]
Associate Professor, Dept. of Computer Science & Engineering, LIET, Hyderabad[2]
AcademicStudent@gmail.com[1]
AcademicGuide101@gmail.com[2]

--- ***---

Abstract:- With the rapid development of the Internet and other associated technologies, many online platforms have started selling data to their customers so that the customers can make use of the data and get new clients and expand their businesses. Hence the collection of person-specific data has gained much importance in recent times. In the data market scenarios, the raw data is collected from the data contributors and uploaded into the cloud servers. This stage is called the data acquisition layer. The cloud service provider processes the data and shares it with the data consumers as per their requirements. But in real-time scenarios, the data consumer cannot verify whether the data collected and processed is truthful and accurate. The data contributors are collecting and handling sensitive information and hence they might not be ready to share sensitive information with unauthorized users and uploaded to a public system like the cloud. Hence it is very essential to ensure the truthfulness, accuracy, integrity, and privacy of the data is protected while using it in a cloud environment. In this project, we have come up with a novel mechanism that ensures the accuracy integrity privacy, and truthfulness of data in the cloud environment using techniques like encryption, identity-based signature, batch verifications, etc. The model has been tested with the real-time data set and it produced desirable results. The system outperforms the existing system and it is scalable for large data markets. The computation and communication expenses incurred during this process are also low compared to the traditional system.

Keywords:- *Cloud storage, public cloud auditing, secure deduplication, batch verification*

---***---

I INTRODUCTION

With the emergence of big data technology and the growing number of users using online platforms for sharing their personal information many organizations have realized the importance of personal information and are using the personal data requirements uploaded on social networking platforms to gain insights into the preferences of the users and are providing data analysis solutions to E-Commerce platforms so that they can reach wider audiences and expanded their businesses [1]. However, in this process, there is a high chance that the security, privacy, and truthfulness of the data might be compromised. The data consumers might end up paying money for inaccurate data. Hence the accuracy and truthfulness of the data must be protected during the process of data acquisition, data processing, and data trading and it should be verifiable by the consumer [2][8].

While it is important to safeguard the privacy and accuracy of data, it is also important that the cloud service provider makes maximum profit while trading the data and reduce the operational costs that incur during data acquisition [10]. To maximize the profit and minimize the expenditure a collusive data service provider might tamper, manipulate, or introduce fake data into the raw data that has been received during data acquisition[4]. Manipulating the data in this way and selling it to potential customers might have devastating effects on the consumer business rather than being able to expand it. Such collusive behaviors cannot be identified, and pose a high risk to the consumer business. Hence the main problems that must be overcome in real-time data markets address below:

1. The data truthfulness and accuracy should be verifiable by the consumer

2. The truthfulness of data collection should be verifiable [3]
3. The truthfulness of data while data processing should be guaranteed.
4. The data market system should be scalable and it should support many data contributors and consumers to maximize profits [6].

In this project, we have come up with a novel technique to address the above issues by using encryption mechanisms that are partially homomorphic and signature-based identification methods to authenticate the contributors and verify the truthfulness uploaded by the contributors [7][9].

Two real-time datasets have been tested using this model and the scalable up to 1,000,000 data contributors per session. This model has achieved desirable results by safeguarding the truthfulness and accuracy of data.

II LITERATURE SURVEY

1. **A novel privacy-preserving authentication and access control scheme for pervasive computing environments**

   Pervasive computing has two important objectives: Privacy and security. Both are divergent in nature. The cloud service provider, on one hand, would like to maintain information in a secure way requesting the details from users to allow only authentic users to log into the system but on the other hand, privacy is required where the details should not be disclosed. This paper provides an authentic way to maintain both privacy and security by amalgamating so important mechanisms called blind signature and hashchain protocols along with Secret key establishment. These cryptographic techniques together not only allow the users to anonymously interact with the cloud server but also provides mutual authentication between the server and user.
2. **Data markets in the cloud: An opportunity for the database community**
   In recent times go to the rapid growth in cloud-related technologies, cloud computing is set to transform multiple domains and bring out new ways to access the data. This resulted in emerging data markets and data associated services. This paper discussed some opportunities, issues, and resolution strategies concerning database users. A data market is a place or a centralized point where data and associated services are sold. The pricing related aspects of data are still in the early stages and need a lot of brainstorming. The aspects of pricing the data after transformations are yet to be decided.
3. **Anonymous publication of sensitive transactional data**
   Ongoing research on Safeguarding privacy and publishing data emphasizes on relational data. It aims to implement privacy-preserving methodologies such as k-anonymity and l-diversity. They mitigate the data loss that happens during the process of anonymization. The prevailing mechanisms tend to work with schema data that has low dimensions. But in real-time there might be applications whose privacy has to be safeguarded despite having hundreds or thousands of dimensions. In this paper, two classes of new anonymization methods for high dimensionalities are proposed. One method is based on the nearest neighbor method and the second method proposes two data transformations to understand and grasp the relationship between the underlying data. These relationships help to form anonymized groups with low information loss. Real-time datasets have been used to prove our point and the results were satisfactory the comparison of these techniques is detailed out in this paper.

III SYSTEM ANALYSIS

Existing system:

In the existing system, a two-layer model is used in the data market scenario. The two layers are the data acquisition layer and the data trading layer. In the data acquisition layer, the data contributors need to submit high-quality data. But in real time it might be possible that the data contributor might manipulate the data for the sake of incentives. Here the truthfulness of data is being compromised. In the data trading layer, the service provider must process the raw data submitted by the data contributors and sell them today data consumers who named them. In real-time, there is a possibility that the service provider might manipulate the data provided by the data contributors and sell them to the consumers. This results in the consumer paying up money today cloud service provider for inaccurate data. We observed that in both the layers the data truthfulness and privacy are being compromised.

Disadvantages:

1. Data truthfulness is not guaranteed in both layers
2. data privacy is not guaranteed in both layers

Proposed system:

In the proposed system, we consider truthfulness and privacy issues that might pop up in real-time and we propose a model that can handle these issues. In the system, we ensure that the data consumer receives correct and complete data without any tampering or manipulations. The proposed system also ensures that data confidentiality is maintained. In the data acquisition layer, the contributor can log in to the system only after the cloud service provider approves his request. He cannot impersonate as any other user. The cloud service provider approves the request only after the necessary background checks. The contributor is not allowed to tamper with the data as it is collected using tamper-proof devices and uploaded directly to the system. When the data is uploaded into the system it is encrypted and uploaded so that manipulation or tampering does not happen in the data trading layer. In the data trading layer, the consumer can request for data of his interest. The consumer's request must be approved by the cloud service provider. As the data is in encrypted format, the consumer cannot see it without the private key. When the cloud service provider grants permission to the data a private key is generated and a link to view the data is given to this consumer. The cloud service provider himself cannot view the raw data as it is in encrypted format. Hence, he can only process it but not manipulate it. Signature-based identification schemes and encryption techniques are being used in this proposed system to accomplish data privacy, integrity, and truthfulness.

Advantages:

1. Data truthfulness is accomplished.
2. Data privacy and integrity are preserved.

IV IMPLEMENTATION

This project has the below three modules:

1. Data Contributor module
2. cloud service provider module
3. data consumer module

Data Contributor:

In this module, the data contributor first registers on the system. Once he is registered, a request will be sent to the cloud service provider for approval. Once the cloud service provider approves the request the data contributors can log in to the system with his username and password. He then logs in with his username and password and uploads the data without tampering with it. He cannot impersonate as another person as his user ID and password is unique and will not be shared with anyone. The data collected will not be tampered with as it is regarded through some devices and uploaded into the system directly. This ensures data truthfulness at the contributor level.

Cloud service provider:

In this module, the cloud service provider first logs in into the system with his credentials. Once he logs in he can perform various activities that pertain to his role like viewing and approving data contributor registration requests, viewing and approving data consumer registration requests, viewing consumers' search requests, granting permissions to data consumers on their search request, and viewing and processing the raw data. Here the cloud service provider has access to the fields which are necessary for processing the data. He cannot see any other personally identifiable information about the data collected by the data contributors. The data that is uploaded by data contributors is in encrypted format and cannot be viewed by the cloud service provider. This ensures privacy at the cloud service provider level along with truthfulness of data as the data uploaded by data contributor cannot be modified by the service provider.

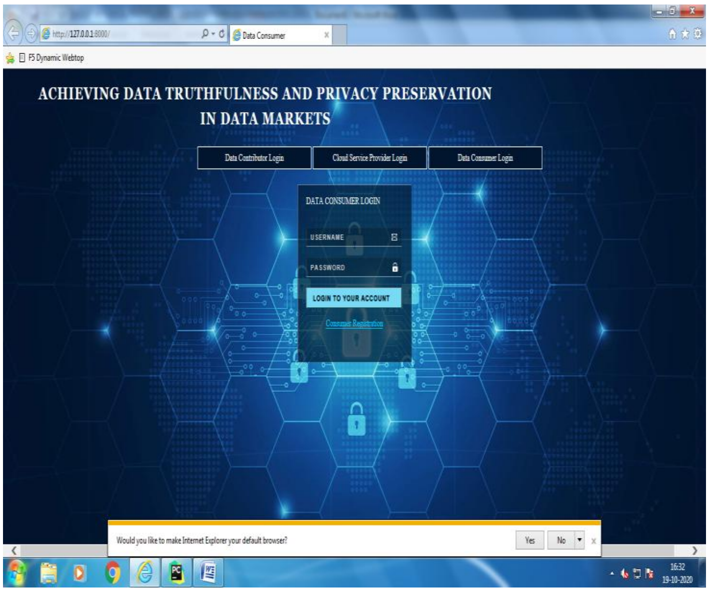
Data consumer:

In this model, the data consumer first registered on the system if he would like to get access to the data for business expansion purposes. Once the data consumer registers him registers himself on the system, this request will be sent to the cloud service provider for approval. Once the cloud service provider approves the registration request, the status of the data consumer would be changed from inactive to active state and he can log in to the system with the username and password that he has set for himself. if the user ID and password combination is correct, he's allowed to log into the system and redirected to a search page where he can search and place a request for the data he would like to have. Once the request is placed, the cloud service provider must grant permission to view the data. Once the cloud service provider grants access, a private key is generated, and enter consumer he's given access to the data. He can view the data once he is permitted to do so. In this way, the consumer can view private and truthful data without any tampering or manipulation during processing at various stages.

V PROJECT EXECUTION AND TESTING

Data Consumer Login page:

The data consumer login page is the default homepage of the application. It has an option to register in case any user is visiting the application for the first time. The user has to first register and then generate his user ID and password. The user will be allowed to log in only if the cloud service provider approves his registration. Once the cloud service provider approves the registration, the user can log in to the website which is username and password. Otherwise, he would get a message that he has to wait as the cloud service provider is yet to approve his registration request.



http://127.0.0.1:8000/

ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

DATA CONSUMER LOGIN

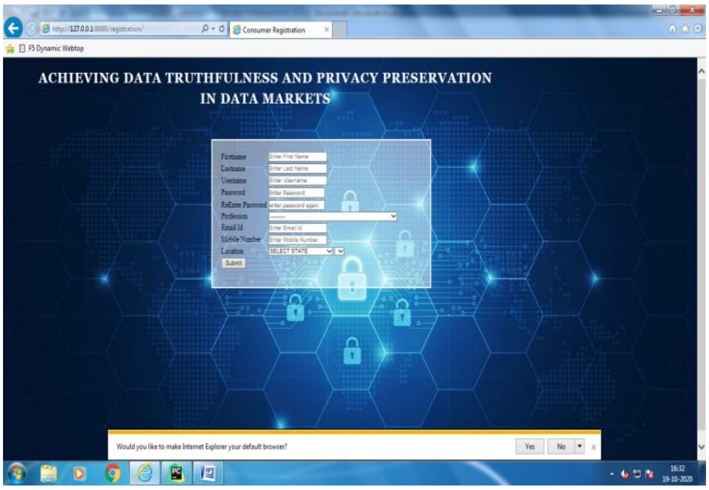
USERNAME

PASSWORD

Would you like to make Internet Explorer your default browser?

Consumer Registration Page:

This page is the user registration page. Any new user who wishes to access the application must first register on this page. The user should save or remember the username and password that he is giving here to log in to the application. Once the user successfully registers his details are saved to the database and he would be waiting for the cloud service provider to approve his request.



http://127.0.0.1:8000/registration/

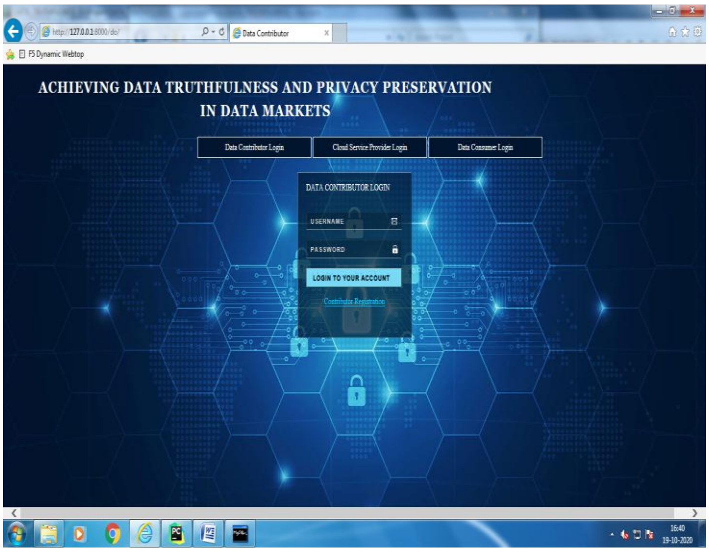
ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Firstname Enter First Name
Lastname Enter Last Name
Username Enter Username
Password Enter Password
ReEnter Password enter password again
Profession -------
Email Id Enter Email Id
Mobile Number Enter Mobile Number
Location SELECT STATE

Would you like to make Internet Explorer your default browser?

Data Contributor Login page:

This page is the data contributor login page. It has an option to register in case any contributor is visiting the application for the first time. The contributor must first register and then generate his user ID and password. The contributor will be allowed to log in only if the cloud service provider approves his registration. Once the cloud service provider approves the registration, the contributor can log in to the website which is username and password. Otherwise, he would get a message that he has to wait as the cloud service provider is yet to approve his registration request.



http://127.0.0.1:8000/do/

ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

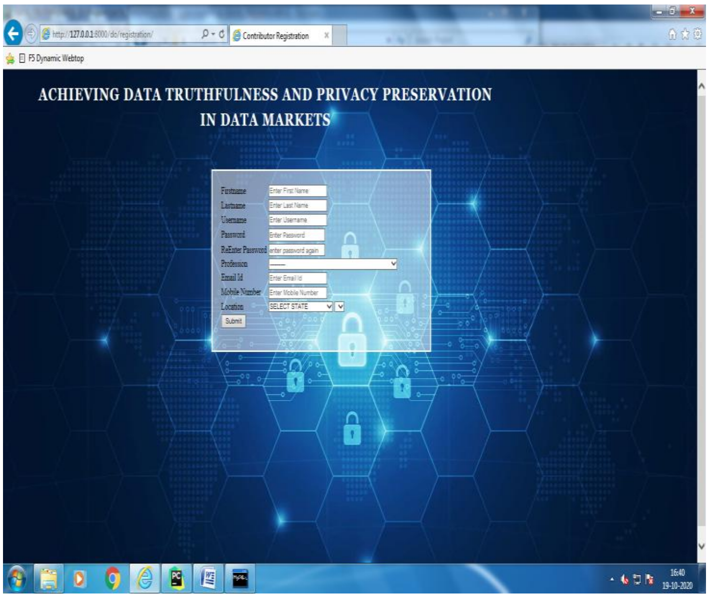
DATA CONTRIBUTOR LOGIN

USERNAME

PASSWORD

Contributor Registration Page:

This page is the contributor registration page. Any new contributor who wishes to access the application must first register on this page. The contributor should save or remember the username and password that he is giving here to log in to the application. Once the user successfully registers his details are saved to the database and he would be waiting for the cloud service provider to approve his request.

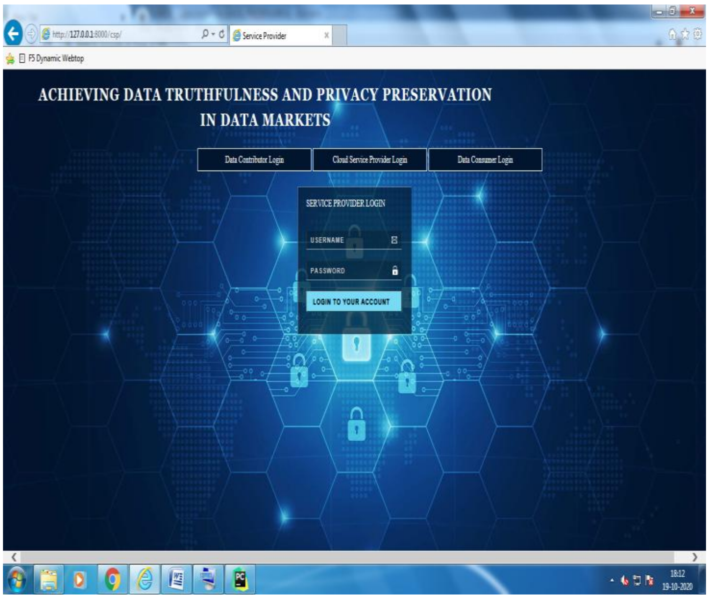


ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Firstname: Enter First Name
Lastname: Enter Last Name
Username: Enter Username
Password: Enter Password
ReEnter Password: enter password again
Profession: ------
Email Id: Enter Email Id
Mobile Number: Enter Mobile Number
Location: SELECT STATE

Cloud service provider login page:

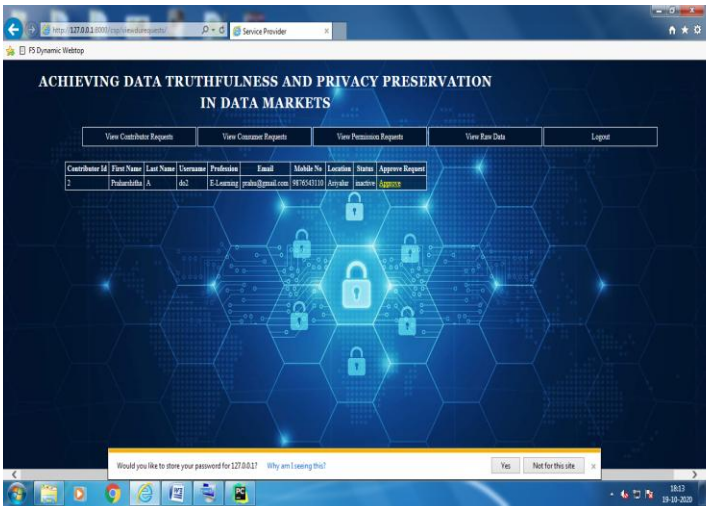
This is the login page for the cloud service provider. He can give his credentials and log in to the platform to perform various functions associated with this role. If the username and password given by the admin are correct, the application allows him to log in and redirect him to the home page. Otherwise, it gives a message that the username and password combination is incorrect.



ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

View contributor requests:

On this page, the cloud service provider can view the registration requests sent by contributors. the requests which are pending for approval and whose status is inactive will be visible on this page. Once the cloud service provider approves the request, the contributor can log in with his credentials into the system.



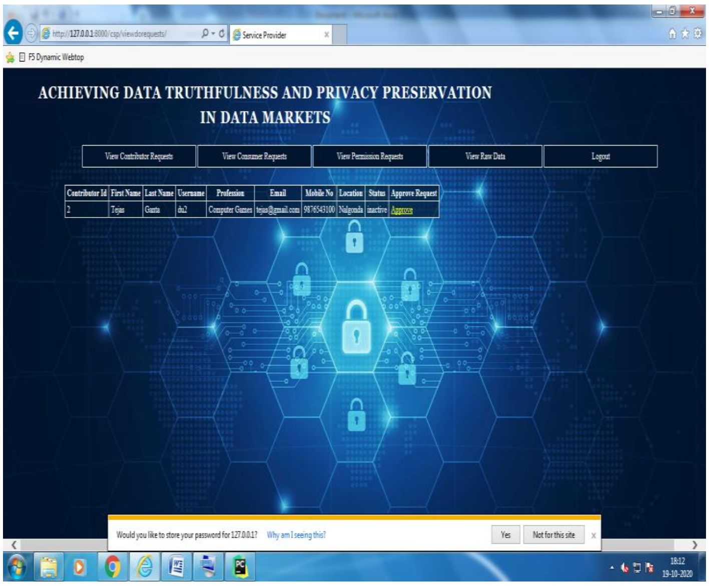
ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Contributor Id	First Name	Last Name	Username	Profession	Email	Mobile No	Location	Status	Approve Request
2	Praharshitha	A	do2	E-Learning	praha@gmail.com	9876543110	Ariyalur	inactive	Approve

Would you like to store your password for 127.0.0.1?

View consumer requests:

On this page, the cloud service provider can view the registration requests sent by consumers. the requests which are pending for approval and whose status is inactive will be visible on this page. Once the cloud service provider approves the request, the consumer can log in with his credentials into the system.



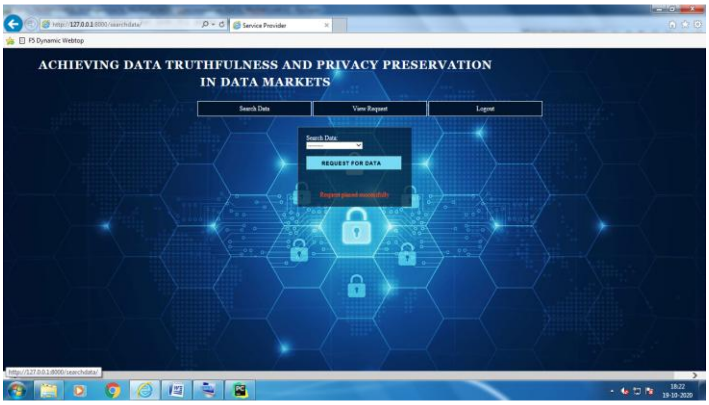
ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Contributor Id	First Name	Last Name	Username	Profession	Email	Mobile No	Location	Status	Approve Request
2	Tejas	Ganta	du2	Computer Games	tejas@gmail.com	9876543100	Nalgonda	inactive	Approve

Would you like to store your password for 127.0.0.1?

. Request data:

If the cloud service provider approves the consumer's registration request, he can give his credentials and log into the system. On successful login, the consumer can search for the data and place a request for the data that he would like to get access to. Once the request is placed the cloud service provider must grant permission to view the data that is there on the cloud.



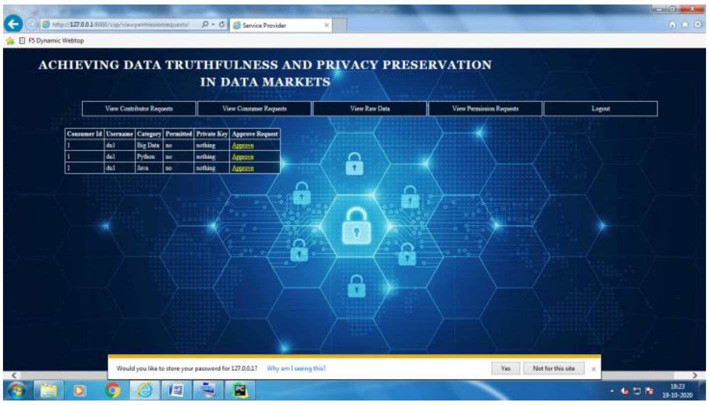
ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Search Data:

Request placed successfully

View permission requests:

Once the consumer places a request for data on a particular topic, The request will be visible in the view permission request page of the cloud service provider. He can choose to approve these requests. Once he approves the requests, a private key would be generated for the consumer and he can view the encrypted data.



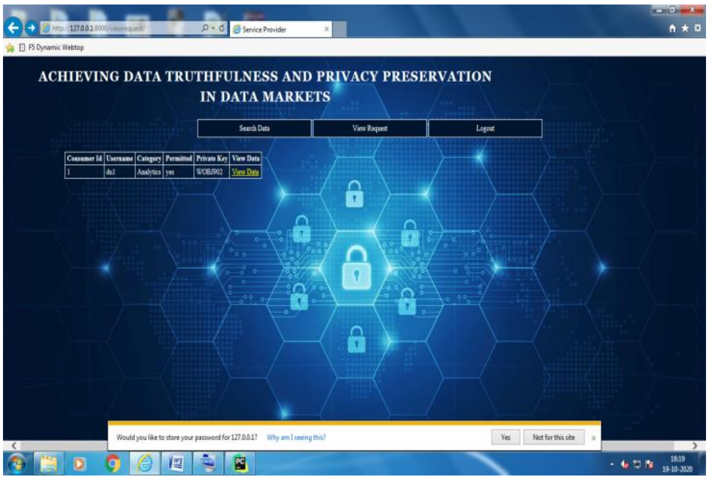
ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Consumer Id	Username	Category	Permitted	Private Key	Approve Request
1	du1	Big Data	no	nothing	Approve
1	du1	Python	no	nothing	Approve
1	du1	Java	no	nothing	Approve

Would you like to store your password for 127.0.0.1? Why am I seeing this?

View Data Request Status:

On this page, the data consumers can view their request status. If the request is approved by the cloud service provider, a private key is generated and they get a link to view the data that they have requested.



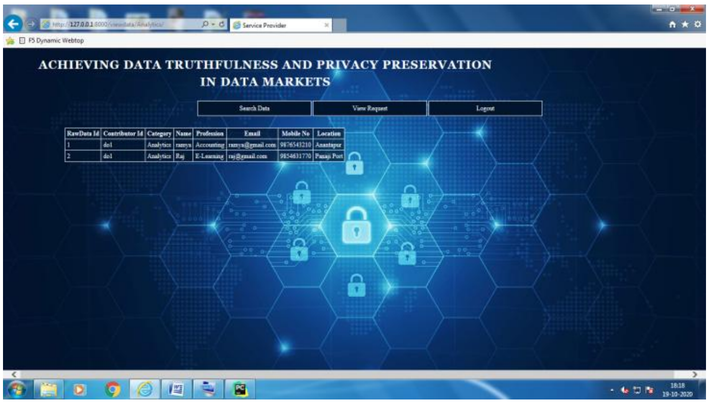
ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

Consumer Id	Username	Category	Permitted	Private Key	View Data
1	du1	Analytics	yes	WOBJ902	View Data

Would you like to store your password for 127.0.0.1? Why am I seeing this?

View Data:

On this page, the data consumers can view the data that they have requested after the cloud service provider has approved their permission requests. If any user tries to view the data without permission he would see encrypted data.

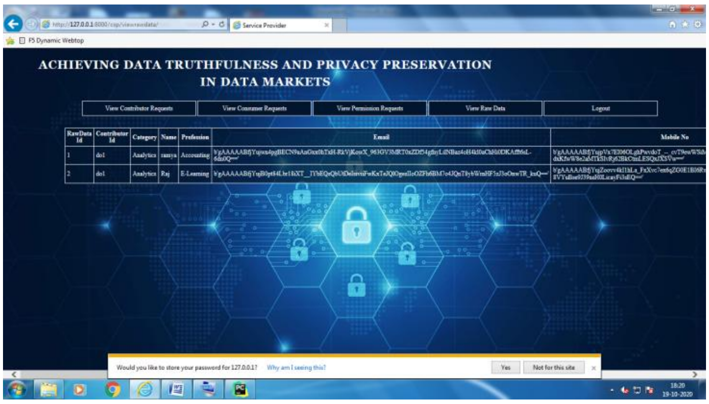


ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

RawData Id	Contributor Id	Category	Name	Profession	Email	Mobile No	Location
1	do1	Analytics	ramya	Accounting	ramya@gmail.com	9876543210	Anantapur
2	do1	Analytics	Raj	E-Learning	raj@gmail.com	9854631770	Panaji Port

View raw data:

We do not give permissions to the cloud service provider the view the data uploaded by data contributors. To protect the truthfulness of the data. The cloud service provider can only view the category and other public details like name and profession. The private data of any user is encrypted and stored in the cloud thus protecting the privacy of the data contributors.



ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

RawData Id	Contributor Id	Category	Name	Profession	Email	Mobile No
1	do1	Analytics	ramya	Accounting	b'gAAAAABfjYujwn4pgBECN9aAnGxn0hTxH-RkVjKoseX_963GV3MRT0xZDf54gfbyLdNBaz4oH4kl0aChHi0DKAff6sL-6dn0Q=='	b'gAAAAABfjYujpVx7E06OLghPwvdoT_...cvT9ewWSd dxKfwW8e2aMTkSIvRj62BkCtmLESQxJXSVu=='
2	do1	Analytics	Raj	E-Learning	b'gAAAAABfjYujB0pt84Lbr18iXT__IYbEQsQbUtDelmviFwKxTaJQ8OpmIloOZFh6BM7o4JQnT8ybWmHF5zJ3oOmwTR_knQ=='	b'gAAAAABfjYujZoovv4kI1hLa_FxXvc7en6qZG0E1BI6R 8VYuBse0J39naH0LicayFi3sEQ=='

Would you like to store your password for 127.0.0.1? Why am I seeing this?

VI CONCLUSION

In this paper, we proposed a novel technique called TPDM which simultaneously safeguards the privacy and truthfulness of data in the data markets where data itself is processed and sold to potential Businessmen who will use it for their business expansion. the one who collects data and uploads it to the online platform it's called a data contributor and he is not supposed to temper the information that he uploads. The cloud service provider who processes the data and sells it should not manipulate or temperature information to make money.

Moreover, the data collected is private information and should not be exposed to unauthorized users. Our technique TPDM satisfies all the above criteria and outperforms existing techniques. This technique has been tested with two real-world datasets and the results were promising. this technique has proved to be scalable for many users and worked well with the semi-honest cloud service provider.

In future, this technique could be expanded to other domains and online platforms which use more complicated Mathematical formula to understand and evaluate its performance in these scenarios.

REFERENCES

[1] P. Upadhyaya, M. Balazinska, and D. Suciu, “Automatic enforcement of data use policies with datalawyer,” in SIGMOD, 2015.

[2] T. Jung, X.-Y. Li, W. Huang, J. Qian, L. Chen, J. Han, J. Hou, and C. Su, “AccountTrade: accountable protocols for big data trading against dishonest consumers,” in INFOCOM, 2017.

[3] G. Ghinita, P. Kalnis, and Y. Tao, “Anonymous publication of sensitive transactional data,” IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 2, pp. 161–174, 2011.

[4] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, “Privacy-preserving data publishing: A survey of recent developments,” ACM Computing Surveys, vol. 42, no. 4, pp. 1–53, Jun. 2010.

[5] R. Ikeda, A. D. Sarma, and J. Widom, “Logical provenance in dataoriented workflows?” in ICDE, 2013.

[6] M. Raya and J. Hubaux, “Securing vehicular ad hoc networks,”Journal of Computer Security, vol. 15, no. 1, pp. 39–68, 2007.

[7] T. W. Chim, S. Yiu, L. C. K. Hui, and V. O. K. Li, “SPECS: secure and privacy enhancing communications schemes for VANETs,” Ad Hoc Networks, vol. 9, no. 2, pp. 189 – 203, 2011.

[8] D. Boneh, E. Goh, and K. Nissim, “Evaluating 2-dnf formulas on ciphertexts,” in TCC, 2005.

[9] R. A. Popa, A. J. Blumberg, H. Balakrishnan, and F. H. Li, “Privacy and accountability for location-based aggregate statistics,” in CCS,2011.

[10] J. H. An, Y. Dodis, and T. Rabin, “On the security of joint signature and encryption,” in EUROCRYPT, 2002.