ACHIEVING DATA TRUTHFULNESS AND PRIVACY PRESERVATION IN DATA MARKETS

A. Durga Devi [1], K. Chandu,

[1]**Assistant professor, PG DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:- adurgadevi760@gmail.com
[2]**PG Student of MCA, Dantuluri** Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:- chandu.k.1315@g.mail.com

ABSTRACT

As a significant business paradigm, many online information platforms have emerged to satisfy society' s needs for person-specific data, where a service provider collects raw data from data contributors, and then offers value-added data services to data consumers. However, in the data trading layer, the data consumers face a pressing problem, i.e., how to verify whether the service provider has truthfully collected and processed data? Furthermore, the data contributors are usually unwilling to reveal their sensitive personal data and real identities to the data consumers. In this paper, we propose TPDM, which efficiently integrates Truthfulness and Privacy preservation in Data Markets. TPDM is structured internally in an Encrypt-then-Sign fashion, using partially homomorphic encryption and identity-based signature. It simultaneously facilitates batch verification, data processing, and outcome verification, while maintaining identity preservation and data confidentiality. We also instantiate TPDM with a profile matching service and a data distribution service, and extensively evaluate their performances on Yahoo! Music ratings dataset and 2009 RECS dataset, respectively. Our analysis and evaluation results reveal that TPDM achieves several desirable properties, while incurring low computation and communication overheads when supporting large-scale data markets.

1 INTRODUCTION

In the era of big data, society has developed an insatiable appetite for sharing personal data. Realizing the potential of personal data�s economic value in decision making and user experience enhancement, several open information platforms have emerged to enable person-specific data to be exchanged on the Internet [1], [2], [3], [4], [5]. For example, Gnip, which is Twitter�s enterprise API platform, collects social media data from Twitter users, mines deep insights into customized audiences, and provides data analysis solutions to more than 95% of the Fortune 500 [2]. However, there exists a critical security problem in these market-based platforms, i.e., it is difficult to guarantee the truthfulness in terms of data collection and data processing, especially when privacies of the data contributors are needed to be preserved. Let�s examine the role of a pollster in the presidential election as follows. As a reliable source of intelligence, the Gallup Poll [6] uses impeccable data to assist presidential candidates in identifying and monitoring economic and behavioral indicators. In this scenario, simultaneously ensuring truthfulness and preserving privacy require the Gallup Poll to convince the presidential candidates that those indicators are derived from live interviews without leaking any interviewer�s real identity (e.g., social security number) or the content of her interview. **Literature Survey**

Power-law distribution of the World Wide Web

Propose an improved version of the Erdös-Rényi (ER) the-ory of random networks to account for the scaling properties of a number of systems, including the link structure of the World Wide Web (WWW). The theory they present, however, is inconsistent with empirically ob-served properties of the Web link structure. Barabási and Albert write that because "of the preferential attachment, a vertex that acquires more connections than anoth-er one will increase its connectivity at a higher rate; thus, an initial difference in the connectivity between two vertices will in-crease further as the network grows. . . . Thus older . . . vertices increase their con-nectivity at the expense of the younger . . . ones, leading over time to some vertices that are highly connected, a 'rich-get-rich-er' phenomenon" [figure 2C of (1)]. It is this prediction of the Barabási-Albert (BA) model, however, that renders it unable to account for the power-law distribution of links in the WWW [figure 1B of (1)]. We studied a crawl of 260,000 sites, each one representing a separate domain name. We counted how many links the sites received from other sites

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

To integrate truthfulness and privacy preservation in a practical data market, there are four major challenges. The first and the thorniest design challenge is that verifying the truthfulness of data collection and preserving the privacy seem to be contradictory objectives. Ensuring the truth fullness of data collection allows the data consumers to verify the validities of data contributors� identities and the content of raw data.

Disadvantages:

Verification in digital signature schemes requires the knowledge of raw data, and can easily leak a data contributor�s real identity. Regarding a message authentication code (MAC), the data contributors and the data consumers need to agree on a shared secret key, which is unpractical in data markets.

Proposed System & alogirtham

We propose TPDM, which achieves both Truthfulness and Privacy preservation in Data Markets. TPDM first exploits partially homomorphic encryption to construct a ciphertext space, which enables the service provider to launch data services and the data consumers to verify the correctness and completeness of data processing results, while maintaining data confidentiality. In contrast to classical digital signature schemes, which are operated over plaintexts, our new identity-based signature scheme is conducted in the ciphertext space.

4.1 Advantages:

- To the best of our knowledge, TPDM is the first secure mechanism for data markets achieving both data truthfulness and privacy preservation.
- TPDM is structured internally in a way of Encrypt then-Sign using partially homomorphic encryption and identity-based signature. It enforces the service provider to truthfully collect and to process real data.

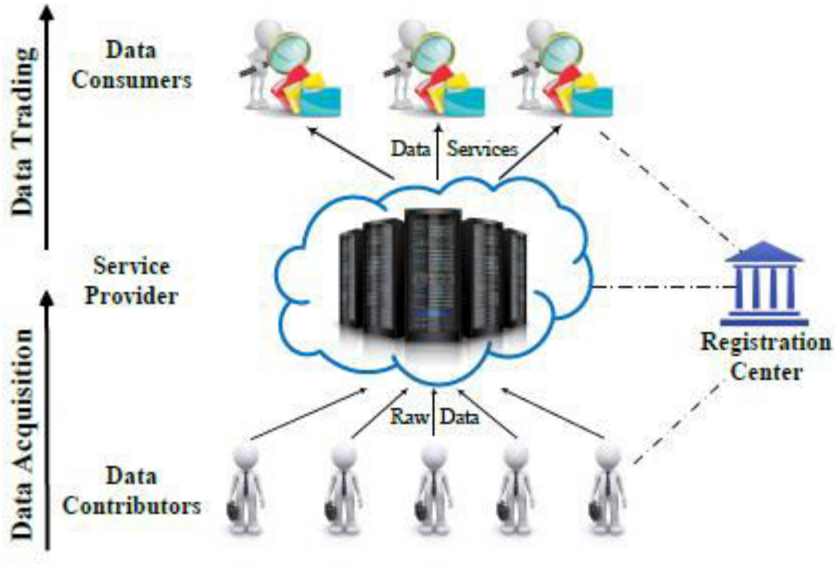


Fig:3.1 System Architecture

IMPLEMENTATION

Introduction of Technologies Used

Initially Java language was called as �oak� but it was renamed as �java� in 1995.The primary motivation of this language was the need for a platform-independent i.e. architecture neutral language that could be used to create software to be embedded in various consumer electronic devices.

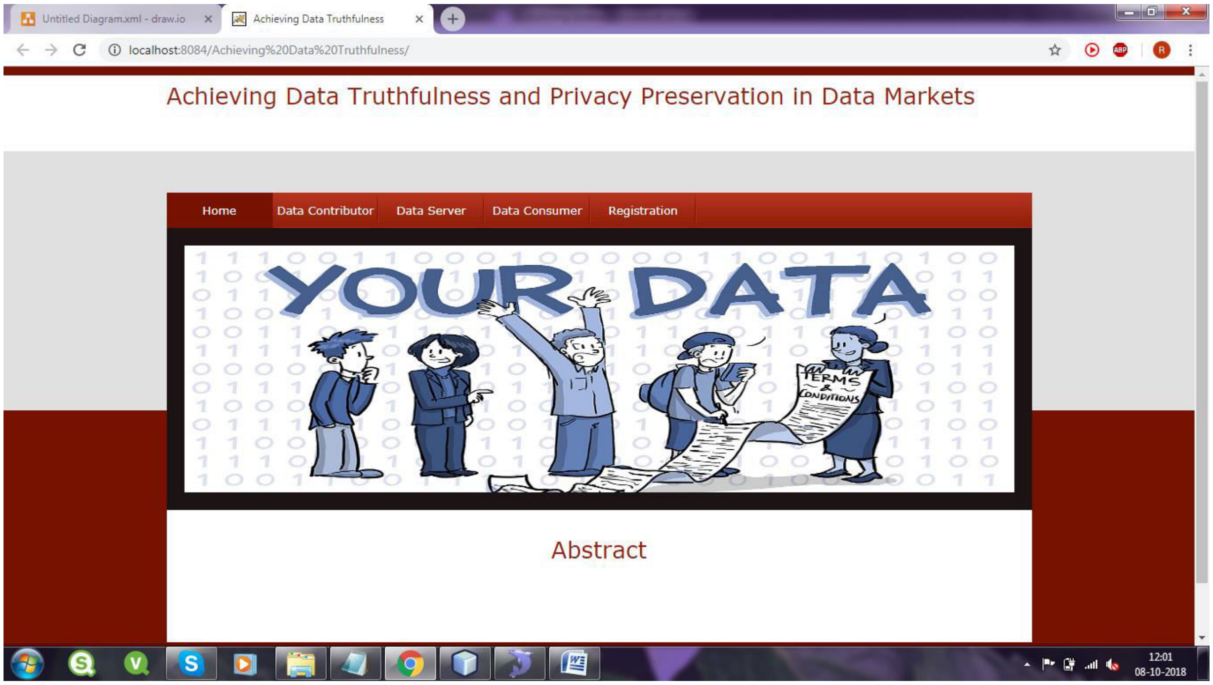
Applications and applets

An application is a program that runs on our Computer under the operating system of that computer. It is more or less like one creating using C or C++ .Java�s ability to create Applets makes it important. An Applet I san application, designed to be transmitted over the Internet and executed by a Java-compatible web browser. An applet I actually a tiny Java program, dynamically downloaded across the network, just like an image. But the difference is, it is an intelligent program, not just a media file. It can be react to the user input and dynamically change.

5 RESULTS AND DISCUSSION

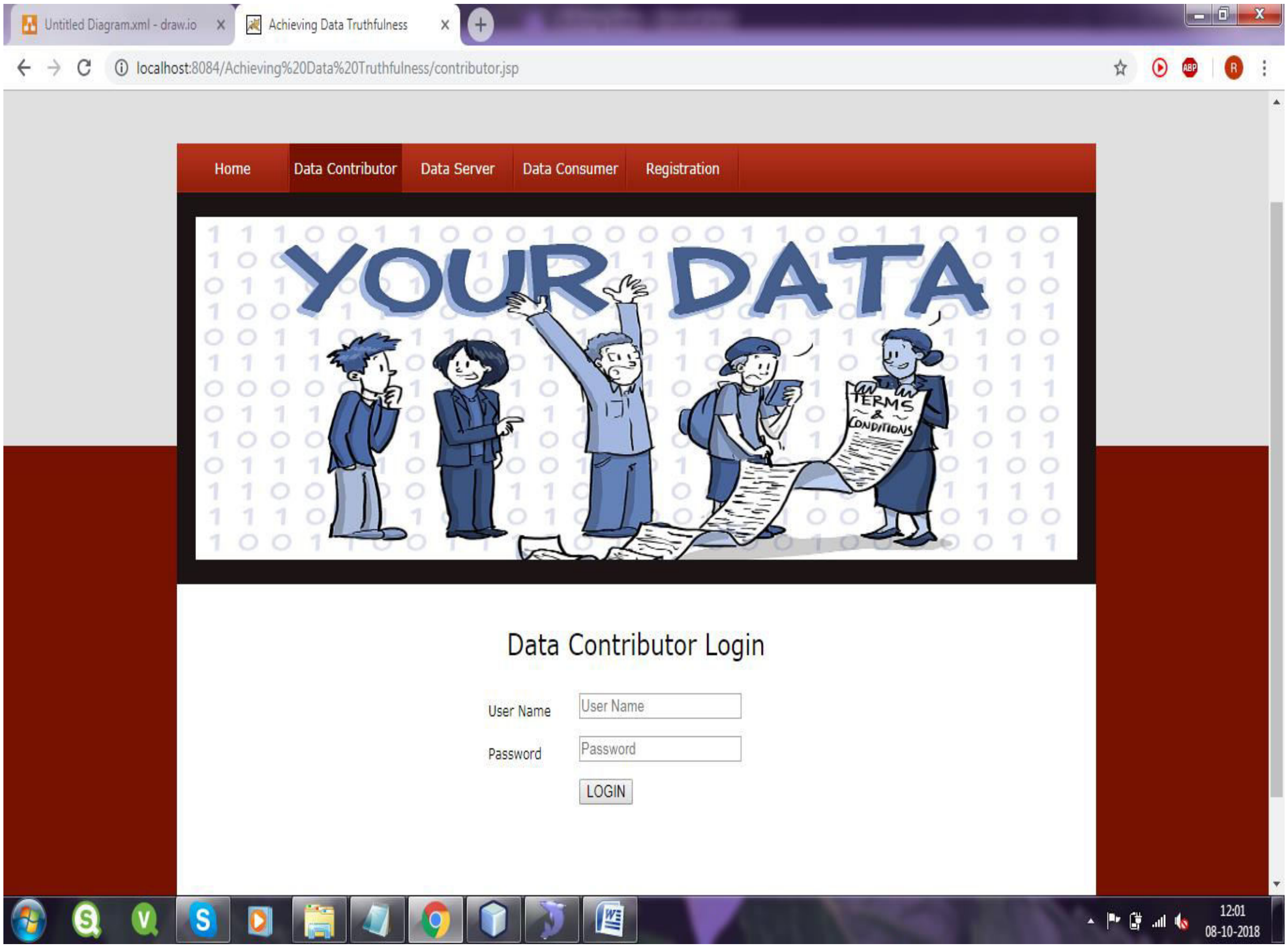
SCREEN SHOTS

Home


Untitled Diagram.xml - draw.io
Achieving Data Truthfulness
localhost:8084/Achieving%20Data%20Truthfulness/
Achieving Data Truthfulness and Privacy Preservation in Data Markets
Home
Data Contributor
Data Server
Data Consumer
Registration
YOUR DATA
TERMS & CONDITIONS
Abstract
12:01
08-10-2018

4

Data Contributor Login:



Untitled Diagram.xml - draw.io
Achieving Data Truthfulness
localhost:8084/Achieving%20Data%20Truthfulness/contributor.jsp
Home
Data Contributor
Data Server
Data Consumer
Registration
YOUR DATA
TERMS & CONDITIONS

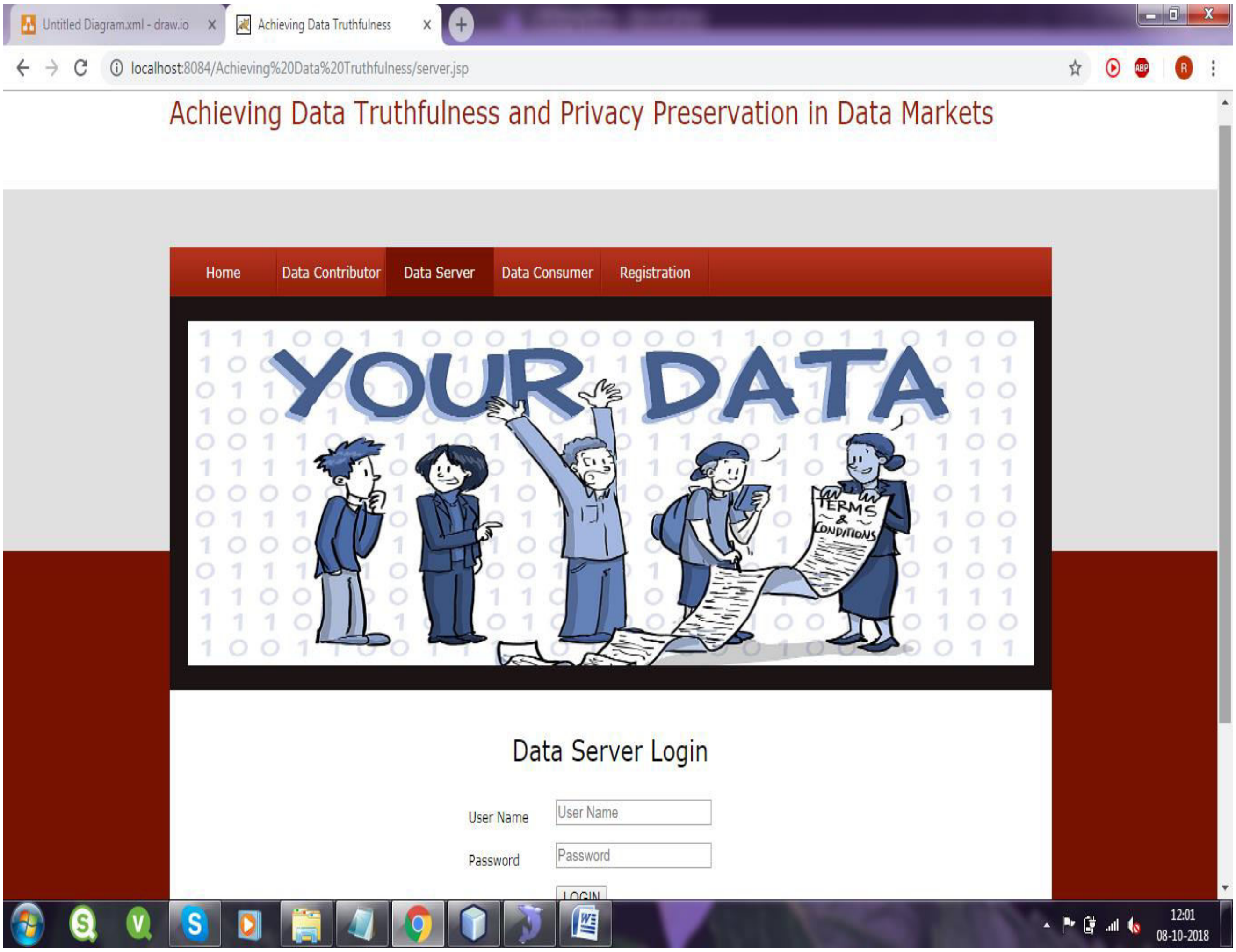
Data Contributor Login

User Name: User Name

Password: Password

LOGIN

Data Server Login:


Untitled Diagram.xml - draw.io
Achieving Data Truthfulness
localhost:8084/Achieving%20Data%20Truthfulness/server.jsp
Home
Data Contributor
Data Server
Data Consumer
Registration
YOUR DATA
TERMS & CONDITIONS
LOGIN
12:01
08-10-2018

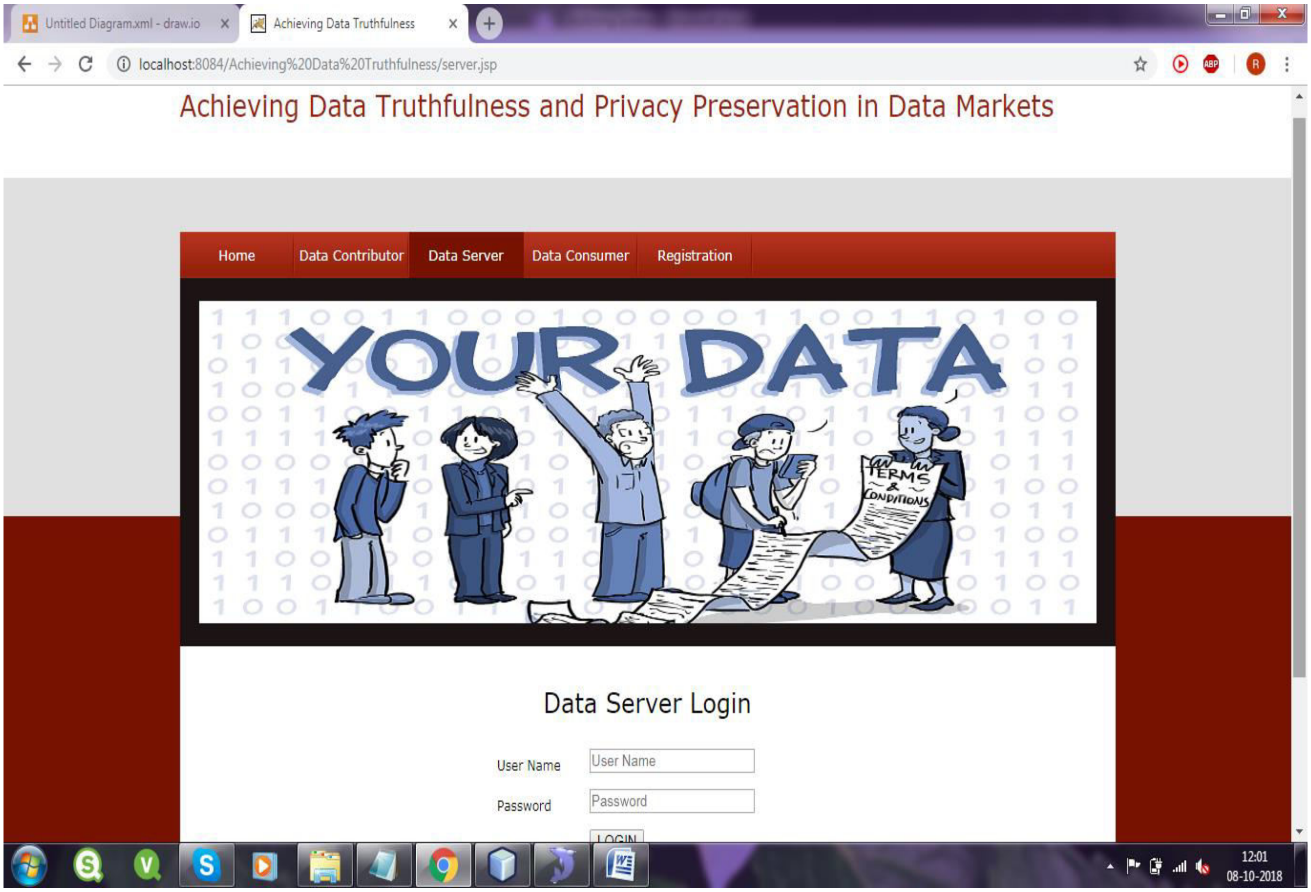
Achieving Data Truthfulness and Privacy Preservation in Data Markets

Data Server Login

User Name User Name

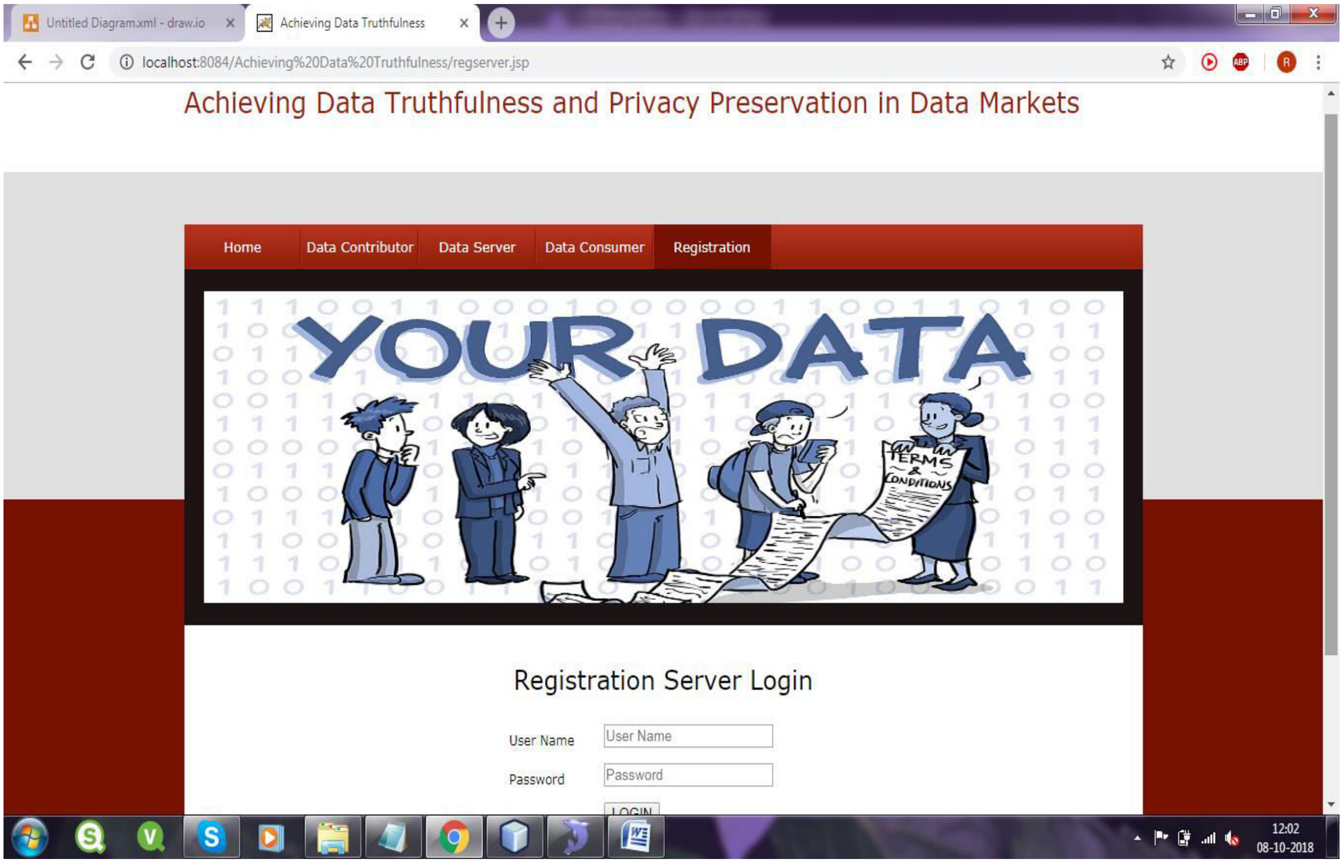
Password Password

Data Consumer Login:



Untitled Diagram.xml - draw.io
Achieving Data Truthfulness
localhost:8084/Achieving%20Data%20Truthfulness/server.jsp
Achieving Data Truthfulness and Privacy Preservation in Data Markets
Home
Data Contributor
Data Server
Data Consumer
Registration
YOUR DATA
TERMS & CONDITIONS
Data Server Login
User Name
Password
LOGIN
12:01
08-10-2018

Registration Server:



Untitled Diagram.xml - draw.io
Achieving Data Truthfulness
localhost:8084/Achieving%20Data%20Truthfulness/regserver.jsp
Home
Data Contributor
Data Server
Data Consumer
Registration
YOUR DATA
TERMS & CONDITIONS
LOGIN
12:02
08-10-2018

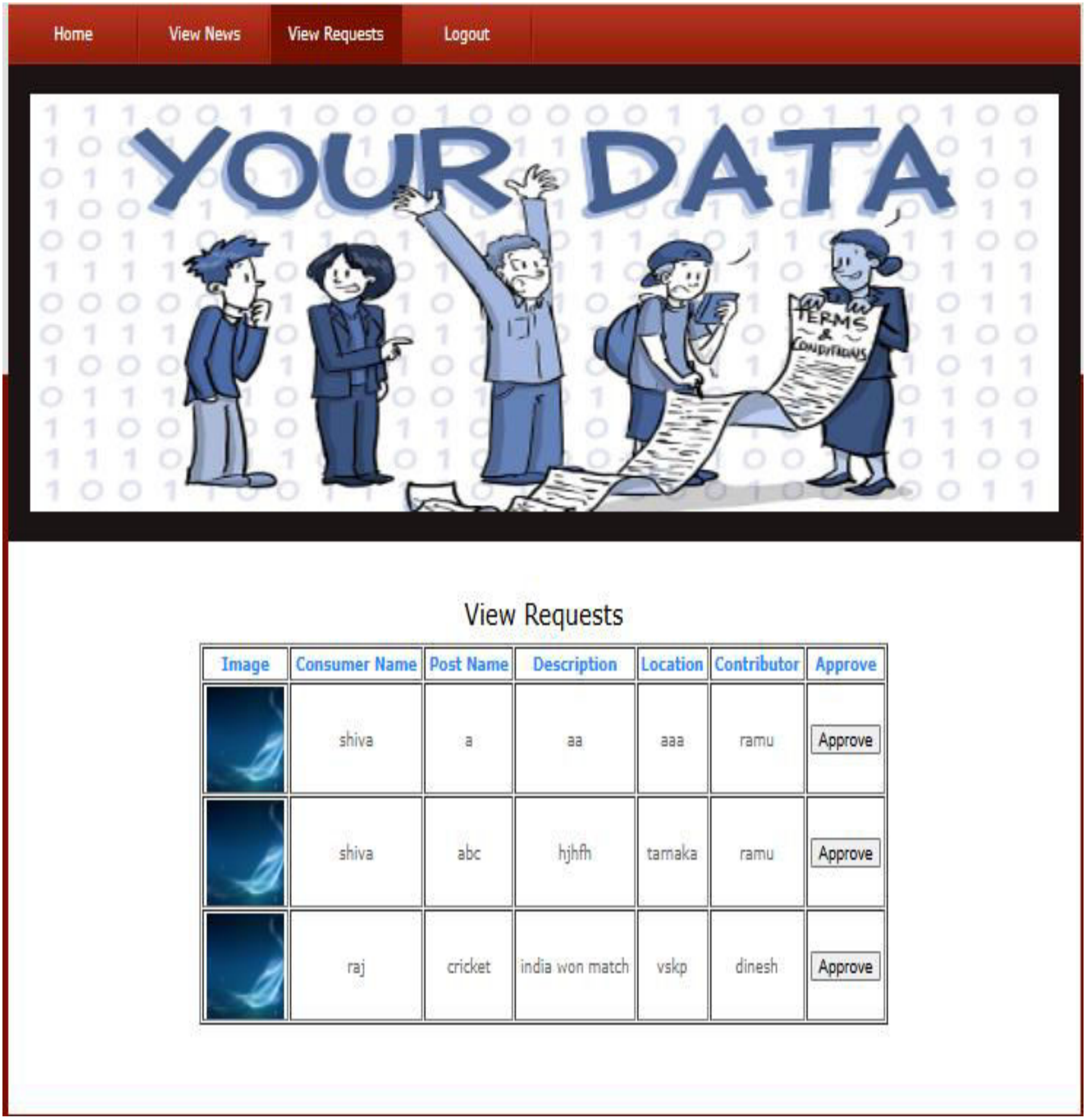
Achieving Data Truthfulness and Privacy Preservation in Data Markets

Registration Server Login

User Name: User Name

Password: Password

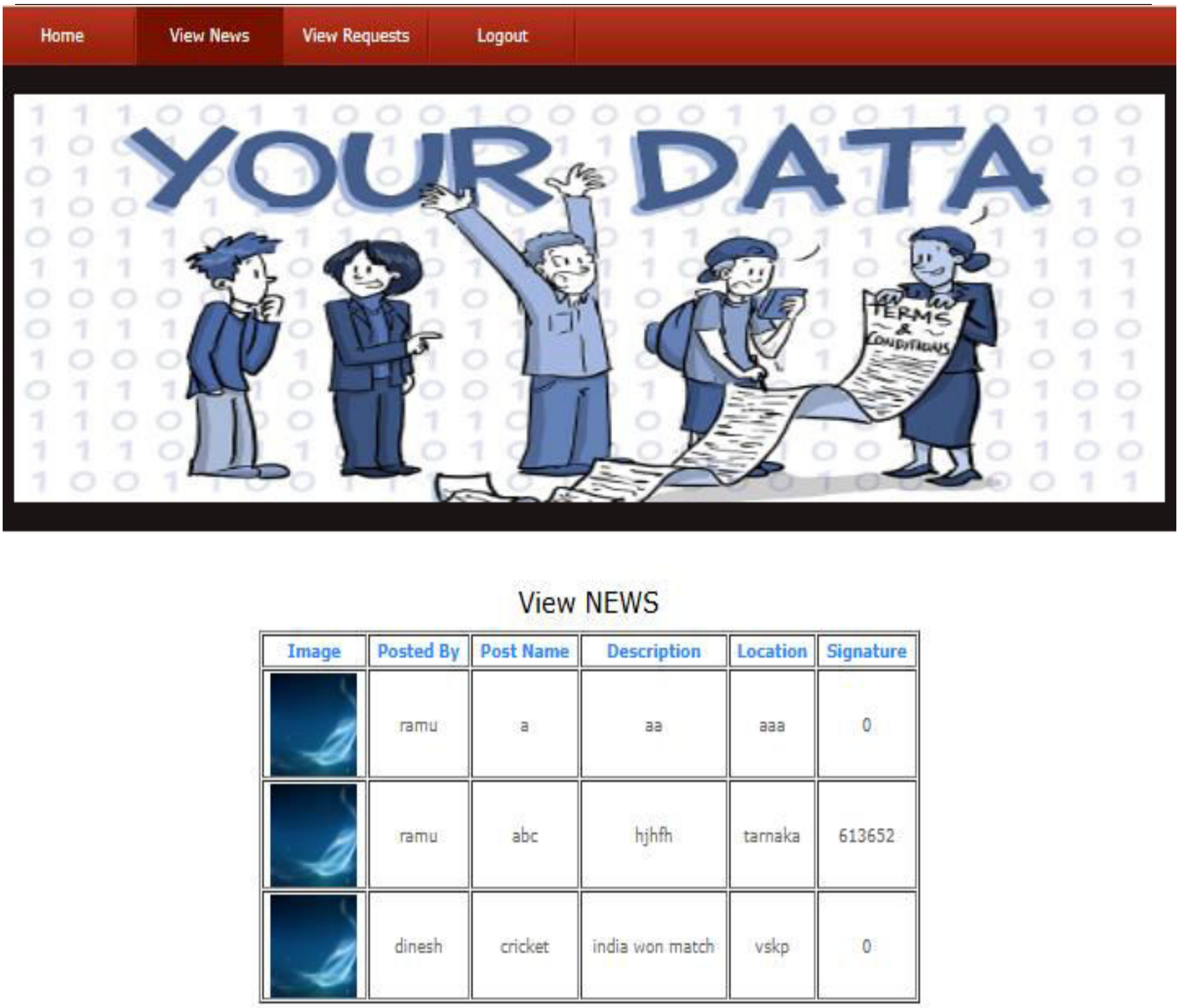
VIEW REQUEST:



View Requests

Image	Consumer Name	Post Name	Description	Location	Contributor	Approve
	shiva	a	aa	aaa	ramu	Approve
	shiva	abc	hjhfh	tamaka	ramu	Approve
	raj	cricket	india won match	vskp	dinesh	Approve

VIEW NEWS:



View NEWS

Image	Posted By	Post Name	Description	Location	Signature
	ramu	a	aa	aaa	0
	ramu	abc	hjhfh	tarnaka	613652
	dinesh	cricket	india won match	vskp	0

6. CONCLUSION AND FUTURE WORK

CONCLUSION

In this project, we have proposed the first efficient secure scheme TPDM for data markets, which simultaneously guarantees data truthfulness and privacy preservation. In TPDM, the data contributors have to truthfully submit their own data, but cannot impersonate others. Besides, the service provider is enforced to truthfully collect and process data. Furthermore, both the personally identifiable information and the sensitive raw data of data contributors are well protected. In addition, we have instantiated TPDM with two different data services, and extensively evaluated their performances on two real-world datasets. Evaluation results have demonstrated the scalability of TPDM in the context of large user base, especially from computation and communication overheads. At last, we have shown the feasibility of introducing the semi-honest registration center with detailed theoretical analysis and substantial evaluations. As for further work in data markets, it would be interesting to consider diverse data services with more complex mathematic formulas, e.g., Machine Learning as a Service (MLaaS) [25], [45], [46]. Under a specific data service

7. REFRENCES

[1] Microsoft Azure Marketplace, https://datamarket:azure:com/ home/.

[2] Gnip, https://gnip:com/.

[3] DataSift, http://datasift:com/.

[4] Datacoup, https://datacoup:com/.

[5] Citizenme, https://www:citizenme:com/.

[6] Gallup Poll, http://www:gallup:com/.

[7] M. Barbaro, T. Zeller, and S. Hansell, A face is exposed for AOL searcher no. 4417749, New York Times, Aug. 2006.

[8] 2016 TRUSTe/NCSA Consumer Privacy Infographic - US Edition, https://www:truste:com/resources/privacy-research/ ncsa-consumer-privacy-index-us/.

[9] K. Ren, W. Lou, K. Kim, and R. Deng, A novel privacy preserving authentication and access control scheme for pervasive computing environments, IEEE Transactions on Vehicular Technology, vol. 55, no. 4, pp. 13731384, 2006.

[10] M. Balazinska, B. Howe, and D. Suciu, Data markets in the cloud: An opportunity for the database community, PVLDB, vol. 4, no. 12, pp. 14821485, 2011.

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

[12] 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.

[13] 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. 161174, 2011.

[14] 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. 153, Jun. 2010.

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

[16] M. Raya and J. Hubaux, Securing vehicular ad hoc networks, Journal of Computer Security, vol. 15, no. 1, pp. 3968, 2007.