IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DATA PRIVACY PRESERVING AND VERIFICATION TECHNIQUES FOR CLOUD COMPUTING SYSTEMS

MOHD UZAIR ILHAJUDDIN FAROOQUI[1], ASST.PROF. V. S. KARWANDE[2]

ME Student, Department of Computer Science & Engineering, EESGOI , India[1]

HOD, Assistant Professor, Department of Computer Science and Engineering, EESGOI,India. [2]

Abstract: In response to society's need for personal data, a number of online information platforms have evolved as a crucial business model. In these platforms, a service provider purchases raw data from data suppliers and subsequently provides value-added services to data customers. The data trading layer presents a significant challenge for data consumers, too, namely how to confirm that the service provider is indeed obtaining and processing their data. Furthermore, data contributors often are reluctant to provide sensitive personal information to data consumers or to reveal their true identity. A method for identifying contributors that is based on honesty is given. A user can submit an assessment to the system after purchasing a product to see if contributors are permitted. Data contributors cannot personally identify others under Data Verification and Privacy Preservation, even though they must legally provide their own data. Furthermore, it is required of the service provider to accurately gather and manage data. Furthermore, critical raw data and the personally identifiable information of data providers are appropriately safeguarded. For data verification and privacy protection, we also used two more data services, and we closely evaluated their effectiveness on two real-world datasets.

Keywords: Emergency demand response (EDR); Advanced Encryption Standard(AES);Access control (AC); Fair and Privacy-Preserving Deep Learning (FPPDL).

I INTRODUCTION

The Data verification and privacy preservation are the goals of the program. It is the first secure data crowdsourcing platform that guarantees privacy and data accuracy. Integrating honesty and privacy into a useful data crowdsourcing system faces four primary challenges. The first and most difficult design issue is that privacy and data collection accuracy validation appear to be at odds with one another. By verifying that data collecting is honest, data consumers can verify the authenticity of data contributor names and raw data contents; however, privacy protecting prevents them from identifying these sensitive contents. Specifically, the attribute of non-repudiation in traditional digital signatures suggests that the signature is irreversible, and any third party can use the data submitter's public key and associated digital certificate to confirm the submitter's legitimacy - that is, the actual purpose of the data collection in our model. However, digital signature systems need raw data to be validated, which might easily reveal a contributor's true identity.

To the best of our knowledge, data verification and protection represent the first safe method in data crowdsourcing systems for both data dependability and preservation. Next, encrypt the sign using a combination of identity-based signature and partially holomorphic encryption for internally organized data verification and privacy protection.

The collection and processing of trustworthy data is required by the service provider. Furthermore, data verification and privacy conservation combines an effective method for results verification that significantly cuts down on

computation time with a two-layer strategy for batch verification. Under Data Verification and Privacy Protection, data contributors are legally required to provide their own data; they are not permitted to replicate other data. Furthermore, it is required of the service provider to accurately gather and manage data. Furthermore, critical raw data and the personally identifiable information of the data suppliers are appropriately safeguarded. In order to preserve privacy and verify data, we have additionally used two more data services, and we have thoroughly examined their effectiveness using two sets of genuine data. The evaluation's conclusions demonstrated the scalability of privacy and data verification and privacy maintenance in the establishment of a broad user base, in particular from computing and communication via heads.

II LITERATURE SURVEY

In this research, The current standalone deep learning system often leads to overfitting and low usefulness. This issue can be resolved by either a distributed system that employs a parameter server to collect local model updates or a centralised framework that trains a global model on data from all parties. Solutions that rely on servers are susceptible to the issue of a single point of failure. In this sense, collaborative learning systems like federated learning (FL) are more resilient. Fairness is a crucial aspect of participation that is overlooked by current federated learning approaches. Every stakeholder receives the same end model, regardless of what they have contributed. To address these issues, we suggest a decentralized Fair and Privacy-Preserving Deep Learning (FPPDL) architecture that incorporates fairness into federated deep learning models. To guarantee correctness and privacy, we design a three-layer onion-style encryption system and a local credibility mutual evaluation system. In contrast to the existing FL paradigm, with FPPDL, every participant obtains a customized FL model whose performance is correlated with his contributions. Benchmark dataset experiments demonstrate that FPPDL achieves an optimal trade-off between accuracy, privacy, and fairness. It makes it possible for federated learning ecosystems to recognize and isolate low-contribution users, enabling more conscientious engagement. [1].

In this research, with the advancement of the Internet of Things (IoT) generation, the value of vast amounts of sensing data will gradually become apparent. Crowd-sourced data trading has so recently attracted a lot of attention as a new business concept. A traditional data trading system consists of a platform, data consumers, and crowd workers. The platform employs crowd laborers to gather data, which it subsequently offers for sale to clients. In this study, we present the DPDT, a differentially private crowd-sensed data trading mechanism that shields task privacy and customer identification from crowd workers as they collect data. DPDT is composed of a differentially private data pricing algorithm and a differentially private data collection technique. The algorithm for data pricing almost perfectly predicts maximum income. Lastly, Finally, thorough simulations are run to demonstrate the DPDT's significant performance. [2].

In this research, Social media data trading has attracted a lot of attention over time. Trading online surfing history in particular is anticipated to bring significant economic benefits to data users, especially when used to tailor advertising. But even with anonymous datasets, user privacy is seriously threatened by the disclosure of whole surfing histories. While a number of current systems investigated the possibility of outsourcing social media data while maintaining privacy, they did not take the impact on the usefulness of the data consumer into account. In this paper, they propose PEATSE, a novel Privacy-preserving data Trading system for online browsing histories. It takes into account customers' different privacy settings and the value of their past online browsing activity. In order to maintain user privacy while weighing the trade-off between privacy and utility, PEATSE modifies users' exact surfing timings on revealed browsing data. Our study and assessment findings based on real-data based trials show that PEATSE accomplishes the following: budget balance, honesty, individual rationality, and protection of user privacy as well as the data consumer's demand for accuracy. [3].

In this research, More often than not, adequate training data is required to train deeply learned models. However, the expensive human process of classifying large numbers of pictures (i.e., annotation) inevitably limits the quantity of real data (training data) that is accessible. To produce extra data for training a deep network, the Generative Adversarial Network (GAN) may be utilized to create artificial sample data (i.e., produced data). However, the generated data is often lacking in annotation labels. To address this issue, we propose a virtual label in this study called Multi-pseudo Regularized Label (MpRL) and apply it to the produced data. The virtual label described in this work is different from traditional labels, which are often a single integral number. Instead, it is a set of weight-based values, each of which is a number in the range of 0 to 1, and it indicates the degree of association between each produced data set and each pre-defined class of actual data. We performed a thorough study of two cutting-edge convolutional neural networks (CNNs) in our testing to determine the efficacy of MpRL. Tests demonstrate that adding MpRL to produced data enhances the performance of individuals in five distinct re-identification datasets. The suggested solution outperforms state-of-the-art techniques on the five datasets, improving rank 1 accuracy by 6.29 percent, 6.30 percent, 5.58 percent, 5.84 percent, and 3.48 percent over a strong CNN baseline. [4].

In this research, Big data is frequently viewed as the secret to releasing the upcoming massive productivity improvement waves. The amount of data collected in our world has been growing as a result of several new applications and technologies that are a part of our everyday life, such as social networking and mobile apps and Internet of Things-based smart-world systems (smart grid, smart transit, smart cities, and so on). Making the most of data has become a major problem due to its exponential expansion. To enable effective data trading, a sizable data market must be established. By releasing data as a commodity into a digital market and encouraging exchange and increased use, data owners and consumers may establish connections with one another data.

However, in order to create a market for data trading that is this successful, a number of issues must be resolved. These issues include figuring out how much is a fair price for the data that is being bought or sold, creating trading platforms and schemes that maximize the social welfare of trading participants while upholding efficiency and privacy, and preventing the traded data from being resold in order to preserve its value. In this study, we conduct an extensive survey on the lifespan of data and data commerce. More precisely, we examine many data pricing models, categorize them into different categories, and weigh the advantages and disadvantages of each approach. We then concentrate on the design of data trading systems and platforms in order to facilitate effective, safe, and privacy-preserving data trade. Ultimately, we discuss digital copyright protection technologies such as digital copyright identifiers, digital rights management, digital encryption, watermarking, and others, as well as data protection concerns throughout the data trade lifecycle. [5].

In this research, Due to the explosive rise of mobile devices, mobile crowdsourcing has become a key area of research. In order to improve the efficacy and veracity of mobile crowdsourcing systems, this research provides an honest incentive mechanism with location privacy protection. TATP is a proposed improved two-stage auction process that takes privacy sensitivity and trust level into account. Moreover, the k-differential privacy-preserving technique is designed to stop users' location data from being compromised. Empirical studies verify the effectiveness of the proposed incentive scheme. Users can be encouraged to participate in sensing chores while maintaining their privacy thanks to the suggested incentive system that safeguards their location privacy. [6].

III. SYSTEMS ARCHITECTURE

The system model architecture Data integrity and privacy are simultaneously guaranteed by the system model architecture's initial effective, secure approach for crowdsourcing systems in the suggested system. With this method, the user purchases goods after sending a system review that looks at whether the contributors are allowed to do so. Under a particular data service, this technology offers verification and privacy protection.

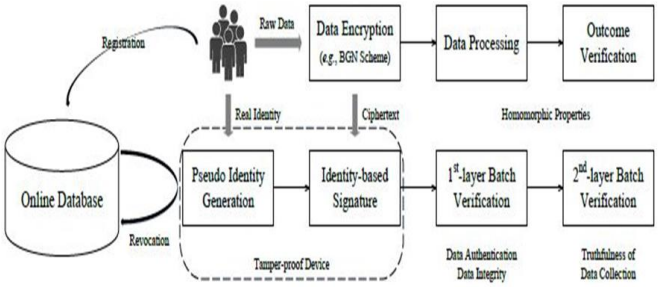

Figure No 3.1: System Architecture

IV EXPERIMENTAL RESULTS

In this section, we show the communication overheads for matching profiles and disseminating data individually. With $\beta = 8$ for the number of random variables, we also show the communication overhead for data distribution in Fig. The graphic illustrates how the service provider's overhead communication increases as the number of data contributors (m) increases. The primary requirement for the service provider is to supply ciphertexts that are linear to m and of the sort $2\beta m$ AES. In contrast, the data consumer's overhead bandwidth stays constant, independent of the data provider, as 2β AES ciphertexts need to be sent for decryption, irrespective of m.

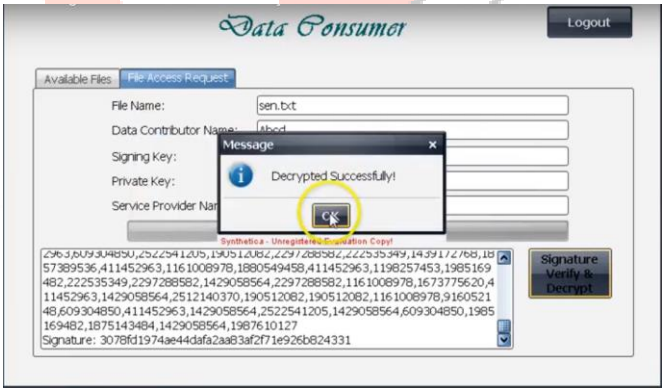

Figure No 4.1: Data Decrypted Process.

V CONCLUSION

Data Crowdsourcing Systems, we have proposed, which involves Data Crowdsourcing Systems, ensures data accuracy and privacy preservation through an effective and secure approach to data verification and protection. Data validation Contributors must provide their own authentic data; they are not allowed to impersonate others. It is also the service provider's responsibility to gather and handle data properly. Furthermore, two alternative data services were

created, and two real-world datasets were used to thoroughly assess each one's performance. Sensitive raw data sources and personally identifiable information are appropriately protected.

REFERENCES

1. X. Ma, Y. Wang, M. E. Houle, S. Zhou, S. M. Erfani, S.-T. Xia, S. Wijewickrema, and J. Bailey, “Dimensionality-driven learning with noisy labels,” ar Xiv preprint arXiv:1806.02612, 2018.

2. Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, Federated Learning. Morgan & Claypool Publishers, 2019.

3. H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, “Fed□erated learning of deep networks using model averaging,” arXiv preprint arXiv:1602.05629, 2016.

4. R. Cummings, V. Gupta, D. Kimpara, and J. Morgenstern, “On the compatibility of privacy and fairness,” 2019.

5. M. Jagielski, M. Kearns, J. Mao, A. Oprea, A. Roth, S. Sharifi Malvajerdi, and J. Ullman, “Differentially private fair learning,” arXiv preprint arXiv:1812.02696, 2018.

6. H. Yu, Z. Liu, Y. Liu, T. Chen, M. Cong, X. Weng, D. Niyato, and Q. Yang, “A fairness-aware incentive scheme for federated learning,” in Proceedings of the 3rd AAAI/ACM Conference on AI, Ethics, and Society (AIES-20), 2020, pp. 393–399.

7. J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. V. Le et al., “Large scale distributed deep networks,” in Advances in neural information processing systems, 2012, pp. 1223–1231.

8. M. Zinkevich, M. Weimer, L. Li, and A. J. Smola, “Parallelized stochastic gradient descent,” in Advances in neural information processing systems, 2010, pp. 2595–2603.

9.Yan Huang, Jingsong Xu, Qiang Wu,Zhedong Zheng,Zhaoxiang Zhang and Jian Zhang,"Multi-pseudo Regularized Label for Generated Data in Person Re-Identification",IEEE Transactions on Image Processing,2018.

10. Fan Liang, Wei Yu, Dou An, Qingyu Yang, Xinwen Fu And Wei Zhao,"A Survey on Big Data Market: Pricing,Trading and Protection",IEEE Access,February 16, 2018.

11.Yingjie Wanga, Zhipeng Cai, Xiangrong Tong, Yang Gao and Guisheng Yin," Truthful incentive mechanism with location privacy-preserving for mobile crowdsourcing systems" Elsevier ,Computer Networks ,2018.

12.Dongqing Liu, Lyes Khoukhi and Abdelhakim Hafid,"Decentralized Data Offloading for Mobile Cloud Computing Based on Game Theory",Second International Conference on Fog and Mobile Edge Computing (FMEC),IEEE,2017.

13.Yutao Jiao, Ping Wang, Dusit Niyato, Mohammad Abu Alsheikh and Shaohan Feng,"Profit Maximization Auction and Data Management in Big Data Markets",IEEE,2017.

14.Chaoyue Niu, Zhenzhe Zheng, Fan Wu, Xiaofeng Gao, and Guihai Chen,"Trading Data in Good Faith: Integrating Truthfulness and Privacy Preservation in Data Markets",33rd International Conference on Data Engineering,IEEE,2017.

15.Song Tan, Debraj De, Wen-Zhan Song,Junjie Yang and Sajal K. Das,"Survey of Security Advances in Smart Grid: A Data Driven

Approach",IEEE Communications Surveys & Tutorials, Vol. 19, No. 1, First Quarter, 2017.

16.Anna L. Buczak and Erhan Guven,"A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection",IEEE Communications Surveys & Tutorials, Vol. 18, No. 2, Second Quarter, 2016.

17.Nguyen Cong Luong, Dinh Thai Hoang,Ping Wang,Dusit Niyato,Dong In Kim and Zhu Han,"Data Collection and Wireless Communication in Internet of Things (IoT) Using Economic Analysis and Pricing Models: A Survey",, IEEE Communications Surveys & Tutorials,2016.