A SURVEY ON DATAVENDING THROUGH CROWD SOURCING BY USING BLOCKCHAIN

[1]Anuja Palhade, [2]Sagar Giri, [3]Omkar Salvi , [4]Satyen Chavan, [5]Mohan Sonawane

[1]Assistant Professor in Department of IT, [2][3][4][5]Research Scholar in Dept. of IT
[1]Dept.Of Information Technology,
[1]Dhole Patil College of Engineering, Pune, India

Abstract: Data Vending is one of the most useful and highly effective concepts that have been utilized to implement data sharing for various different approaches and applications. The Data is one of the most essential commodity that can be highly useful in achieving various machine learning and artificial intelligence applications. But there are certain problems with this approach, due to the fact that the data might contain personally identifiable information or sensitive information. This leads to a lack of trust between the Data vendor and aggregator which can lead to the termination of data sharing. Therefore, there is need for a system that can provide effective security to tampering and increase trust between the data vendor and aggregator. To achieve a system capable of improving the trust in the data sharing paradigm a collection of researches have been analyzed in this research article. This research article has been highly useful in developing our approach which implements the use of the Distributed Blockchain Platform to eliminate the security concerns. The proposed approach will be effectively outlined in the upcoming editions of this research.

Index Terms - *Blockchain, Distributed Systems, Natural Language Processing*

I. INTRODUCTION

The technological advancements have been one of the most important and essential driving forces that have been facilitating an improved lifestyle for the majority of the individuals. The technological advancements are facilitated through the use of an effective and useful research being done. The research can only be performed with effective compilation of the data on the topic and effectively analyzing that data. Over the years there has been numerous researches in various different topics which has led to immense growth and advancement.

This is the age of information, wherein the data or information is one of the most essential commodity that can have a significant impact on society. There are a lot of applications that are dependent on data. Such as machine learning and artificial intelligence which require a large amount of data to be able to extract the effective understanding of the situation or an application.

The machine learning applications can provide a valuable insight by analysis of the data effectively. Therefore, information or data is one of the most effective tool that can be highly useful in a variety of applications. Facebook, twitter and other social media have been one of the largest accumulation of knowledge. Most of these applications are falling short of data for their executions. This is due to the fact that there has been a loss of trust in providing an effective platform for the exchange of data. The data that is being generated is not getting analyzed effectively which is problematic to the platform which has been limiting the effective extraction of valuable insight form the data. The loss of trust is one of the main reasons that can hamper the process of effective and useful data sharing.

To improve the process of data sharing and enhance the trust between the data vendors and aggregators, there is the need for a system that is robust and tamper free. The blockchain platform is one of the most effective techniques to prevent tampering and maintaining the integrity of the data being shared. The distributed platform can be effectively used for the improvement of the security of the approach effectively. Any tampering on the blockchain approach can be identified effectively as there is an avalanche effect. The integrity of the data being shared is maintained through the implementation of the distributed approach.

Therefore, the data can be easily shared by the organization to the data aggregators that will give the data to the interested receiver without any tampering through blockchain distributed framework. The data is encrypted using the RSA approach, and the NLP or Natural Language Processing is realized to achieve the effective processing of the query that is passed by the user.

This literature survey paper dedicates section 2 for analysis of past work as a literature survey, and finally, section 3 concludes the paper with traces of future enhancement.

II RELATED WORKS

D. Dang states crowdsourcing is a new emergent distributed computing and business model on the backdrop of the internet blossoming. It has become a critical challenge to handle worker quality evaluation build on big data and analytics technology. [1] Thus the proposed paper implements a worker quality assessment algorithm that is used for any critical tasks such as tagging, matching, filtering, categorization, and many other emerging applications. The proposed model results to be an effective and accurate demonstrate algorithm. Thus the crowdsourcing is a distributed problem-solving model.

J. Zhou expresses for decentralized applications in recent advances of blockchain technologies have provided exciting opportunities. [2] In recent years most researchers had got attracted to blockchain technology. Blockchain technology is an open and decentralized ledger link of the block and provides proven security promoted by cryptography. Electronic cash system bitcoin had first introduced the blockchain. Thus the interesting topic part of smart contracts facilitate distributed data vending which allow data to transfer securely from one end to the other using blockchain

D. Peng implements the mechanism and calculates its improvement in terms of quality of service and profit of service provider. Applications require manual efforts and physical resources for sensing, processing, and transmitting data in crowdsensing. The quality of crowdsensing is based on its services is faces three major quality estimation, incentive design, and effective feedback. [3] The success of crowdsensing depends on its services critically based on sufficient and reliable data contributions from individual participants. Thus the proposed paper achieves superior performance when compared to the general data collection model and uniform pricing scheme.

K. Yang narrates mobile crowdsourcing network (MCN) is a successful network architecture that implements the principles of crowdsourcing to evaluate tasks with human assessment and powerful mobile devices. In the proposed paper the researcher combined crowdsourcing sensing and crowdsourcing computing a general architecture for a mobile crowdsourcing network. Most mobile devices nowadays are installed with powerful processors, various sensors, large memories, fast wireless communication modules.[4] Thus the proposed paper brings awareness to further investigation on security and privacy solutions for mobile crowdsourcing networks.

R. Ouyang express crowdsourcing is an activity of acquiring required content, information, or services by requesting contributions from a huge group of usually unknown people, rather than from traditional employees or suppliers. It is becoming more and more famous as it issues an easy, time-, and cost-efficient way to collect a huge volume of data for a variety of applications, such as image labeling, image description, sentiment analysis, listing verification, object counting, translation, and logo design. [5] Thus the proposed paper uses new parallel and streaming truth discovery algorithms for quantitative crowdsourcing applications.

A. Azaria describes for electronic medical records (EMRs) bureaucratic inefficiency have slowed innovation for many years. But it creates the problem such as patients to share details of their healthcare and restore agency over their medical data which can be unsafe. [6] Thus to handle this issue the proposed paper has introduced MedRec: a novel, decentralized record management system. Just by leveraging unique blockchain properties, the system gives patients a comprehensive, immutable log and easy access to their medical data across providers and treatment sites

J. Huang states in recent years crowd-sensing systems have achieved considerable interest and adoption. The problem such as privacy leakage and single point of failure was observed in the traditional centralized architecture of crowd-sensing systems. [7] The researcher provides guarantees privacy by providing users to register without a true identity and save encrypted sensory data in the distributed database by using the properties of the blockchain. Security and low service fee have been upgraded in the proposed technique.

J. An introduces a variety of sensors the Internet of Things (IoT) because of the huge number of data that can be collected by the sensors. The data collected by the crowdsensing is varied and is difficult to evaluate. Quality control methods are mostly based on a central platform and the existence of fraud were the two major problems of traditional crowdsensing. To solve these problems of a blockchain is introduced in the proposed paper. [8] To realize the nonrepudiation and no tampering of information credit-based verifier selection mechanism is used in crowdsensing.

H. Duan elaborates on the concept of a new blockchain-powered crowdsensing system. The techniques such as consolidation and customized designs, the framework provides a strong security guarantee with data confidentiality, differential privacy, service correctness, as well as robustness. The system permits data providers to securely contribute data to the transparent blockchain with a confidentiality guarantee on individual data.[9] Thus the proposed experiment results to be a high efficiency of their designs on both mobile client and SGX-enabled server.

J. Xu states Industrial Internet of Things (IIoT) has the prominent potential to gather huge useful information by utilizing the intelligence of crowds or distributed computers. [10] The crowd-intelligence ecosystem involves three stakeholders, namely the platform, workers, and task publisher. The stakeholders may have no mutual trust and their interests conflict. Thus the paper built a distrustful crowd-intelligence ecosystem depend on the common decentralization feature of mobile edge computing and blockchain technology.

A. Lertpiya describes social Listening and chatbots are High-level NLP tasks due to the sheer amount of data to be processed it require assistance from computer systems. [11] Social media is monitored by the Social listening (SL) process for specific terms such as companies, products, or services. In the proposed paper the author used NLP algorithms word segmentation, sentence segmentation, word error detection, word variant detection, named entity recognition. The proposed model is used for evaluating five fundamental Thai NLP tasks on a UGWC corpus.

Dr. N. Srinivasan a larger amount of data is obtained from a large number of smartphone users and further carries out data analysis on the aggregated data. By using their sensor-rich mobile devices the requester known as task owner can crowdsource data from the workers known as smartphone users in mobile crowdsourcing.[12] It has become challenging problems such as data collection, data aggregation, and data analysis for a resource-constrained requester when data volume is extremely large, i.e., big data. The proposed technique greatly reduces the computational cost by using batch verification and data update methods.

III CONCLUSION AND FUTURESCOPE

This research article concerns with the thorough study of various systems for the purpose of data vending. The paradigm of data vending is one of the most effective for enabling easy sharing of data. Data sharing is essential for a variety of applications and fields. This research article aims to identify the various conventional systems that are utilized for the purpose of data vending. This is to identify the inconsistencies and improvements offered by these approaches. After a detailed analysis of the related researches, the distributed blockchain framework has been selected as the system of choice for the Data vending approach. The future editions of the research will elaborate further on the presented technique for the purpose of secure Data Vending through the use of Blockchain.

REFERENCES

[1]   D. Dang, Y. Liu, X. Zhang, and S. Huang, "A Crowdsourcing Worker Quality Evaluation Algorithm on MapReduce for Big Data Applications," in IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 7, pp. 1879-1888, 1 July 2016, doi: 10.1109/TPDS.2015.2457924.

[2]   J. Zhou, F. Tang, H. Zhu, N. Nan, and Z. Zhou, "Distributed Data Vending on Blockchain," 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, NS, Canada, 2018, pp. 1100-1107, doi: 10.1109/Cybermatics_2018.2018.00201.

[3]   D. Peng, F. Wu and G. Chen, "Data Quality Guided Incentive Mechanism Design for Crowdsensing," in IEEE Transactions on Mobile Computing, vol. 17, no. 2, pp. 307-319, 1 Feb. 2018, doi: 10.1109/TMC.2017.2714668.

[4]   K. Yang, K. Zhang, J. Ren, and X. Shen, "Security and privacy in mobile crowdsourcing networks: challenges and opportunities," in IEEE Communications Magazine, vol. 53, no. 8, pp. 75-81, August 2015, doi: 10.1109/MCOM.2015.7180511.

[5]   R. W. Ouyang, L. M. Kaplan, A. Toniolo, M. Srivastava, and T. J. Norman, "Parallel and Streaming Truth Discovery in Large-Scale Quantitative Crowdsourcing," in IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 10, pp. 2984-2997, 1 Oct. 2016, doi: 10.1109/TPDS.2016.2515092.

[6]   A. Azaria, A. Ekblaw, T. Vieira and A. Lippman, "MedRec: Using Blockchain for Medical Data Access and Permission Management," 2016 2nd International Conference on Open and Big Data (OBD), Vienna, 2016, pp. 25-30, doi: 10.1109/OBD.2016.11.

[7]   J. Huang, L. Kong, L. Kong, Z. Liu, Z. Liu, and G. Chen, "Blockchain-based Crowd-sensing System," 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), Shenzhen, 2018, pp. 234-235, doi: 10.1109/HOTICN.2018.8605960.

[8]   J. An, D. Liang, X. Gui, H. Yang, R. Gui and X. He, "Crowdsensing Quality Control and Grading Evaluation Based on a Two-Consensus Blockchain," in IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4711-4718, June 2019, doi: 10.1109/JIOT.2018.2883835.

[9]   H. Duan, Y. Zheng, Y. Du, A. Zhou, C. Wang, and M. H. Au, "Aggregating Crowd Wisdom via Blockchain: A Private, Correct, and Robust Realization," 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, Kyoto, Japan, 2019, pp. 1-10, doi: 10.1109/PERCOM.2019.8767412.

[10]   J. Xu, S. Wang, B. K. Bhargava, and F. Yang, "A Blockchain-Enabled Trustless Crowd-Intelligence Ecosystem on Mobile Edge Computing," in IEEE Transactions on Industrial Informatics, vol. 15, no. 6, pp. 3538-3547, June 2019, doi: 10.1109/TII.2019.2896965.

[11]   A. Lertpiya et al., "A Preliminary Study on Fundamental Thai NLP Tasks for User-generated Web Content," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-8, doi: 10.1109/iSAI-NLP.2018.8692946.

[12]   G. Zhuo, Q. Jia, L. Guo, M. Li, and P. Li, "Privacy-Preserving Verifiable Set Operation in Big Data for Cloud-Assisted Mobile Crowdsourcing," in IEEE Internet of Things Journal, vol. 4, no. 2, pp. 572-582, April 2017, doi: 10.1109/JIOT.2016.2585592.