# ENABLING CROWD SOURCING FOR TASK MANAGEMENT THROUGH BLOCKCHAIN

[1]Sagar Giri , [2]Omkar Salvi , [3]Satyen Chavan , [4] Mohan Sonawane, [5]Anuja Palhade
[1 2 3 4] Research Scholar ,[5] Assistant Professor,
[1]Department of Information Technology,
[1]Dhole Patil College Of Engineering, Pune, India.

***Abstract :*** The paradigm of data vending is one of the most unique approaches that have been the result of increasing evolution and technology leading towards the information age that will live today. Data is highly essential commodity that is required to maintain the steady pace of improvement and technological advancements in this world. This data is crucial for implementations in a wide variety of machine learning implementations and other deep learning evaluations. There is a lack of trust between the various entities in a data vending approach leading to words an effective deterioration of this essential system. Therefore to improve this approach significantly the proposed methodology in this research article elaborate on a data vending approach that achieve effective security through RSA encryption and the distributed Blockchain framework. The proposed methodology has been subjected to rigorous experimentation for its performance evaluation which is achieved significant improvements over the conventional approaches.

***Keywords*** **- Blockchain, Data Vending, Natural Language Processing, RSA, Pearson Correlation.**

## I. INTRODUCTION

Data is considered as one of the most important and highly valuable resources on this planet nowadays. This is due to the fact that data or information can be utilized in a lot of different ways to achieve previously unachievable or gaining valuable insight into particular field or a procedure. There are a lot of different methodologies that benefit from the effective utilization of data analysis approaches which can provide meaningful improvements to the entire field. Therefore data is essential for the purpose of achieving extremely important and useful milestones in the advancement of the human race.

The past few decades and the current one are being considered as the age of information due to vast amount of data and information being available easily. This data is generated in a lot of ways some of it is private data of individuals all social media websites and other web portals and services, along with the data that is generated on E-Commerce websites banking applications which creates massive amounts of data every single day. This data is extremely valuable and needs to be stored and processed effectively.

Most of the business owners utilize this data to achieve analytics that can be crucial for understanding the various nuances of the business. This data can be highly helpful enhancing the model of the business and improving the efficiency of the entire process. For this effective enlightenment there is a need for data processing approaches that need to be utilized to achieve this kind of insight. There are various approaches for handling big data and statistical analysis and evaluation through machine learning methodologies. A lot of businesses that and approaches that provide this kind of data validation service require massive amounts of data to generate accurate and precise predictions or assessment.

But most of the time there is a loss of trust between the data providers as well as the data seekers. This leads to an acute shortage of information that will hinder a lot of the evaluation systems to not be able to achieve precise computations.

This is due to the fact that the data might contain private information or personally identifiable information that will be somehow misused by the data seekers leading to a breach of trust and contract. This is a highly undesirable circumstance which has extremely negative consequences to an effective implementation for data sharing mechanism. This leads to a loss of data vending approaches which are crucial to sustain various machine learning and deep learning implementations.

Therefore there is a need for an effective translate data wedding approach that can allow for effective sharing of data among the data sequence and the data providers through enhanced privacy and security. They have been a large number of researches that have been performed to achieve this data vending approach precisely but a number of them have not been able to achieve a great efficiency or reliability. The four this resort article deals with the elaboration of an effective and useful data wedding approach through the use of the Blockchain platform to achieve highly accurate secure and reliable data vending procedure.

The RSA encryption approach has been utilized in this methodology to achieve effective security of the data and provide a temporary security that is robust in nature. This type of encryption is highly accurate and is asynchronous in nature which provides even better security and reliability. This approach utilizes a clever combination of modulus operations along with the prime and co-prime numbers to achieve accurate and useful encryption. Highly precise reward in printer scheme has also been implemented to achieve increase reliability to the entire procedure.

The proposed paper on distributed data vending dedicates section 2 for the study of the past work under the section Literature Survey. The Details of the proposed system are narrated in the section 3 with the title of the proposed methodology. Obtained results of the system are evaluated using section 4 with the name Result and Discussions and Finally, Section 5 deals with the conclusion and Future scope of the proposed research.

## II. LITERATURE SURVEY

D. Dang explains that the paradigm of crowd sourcing is a highly useful approach for the purpose of achieving effective problem solving capabilities and production requirements for various organizations. This approach has been highly effective in achieving greater efficiency along with reduction of absolute cost of implementation while assuring solutions for various difficulties that are technical in nature. But due to various problems that are faced nowadays for achieving this crowdsourcing model where is businesses have not yet implemented this approach entirely to reap the greater efficiency benefits. The effort to improve this approach and achieve effective improvement the authors in this research article have proposed the use of crowdsourcing in Big data applications along with map reduce for evaluating the quality of work produced by the worker.

J. Zhou expresses that there is increased interest in the Blockchain paradigm in the recent years due to the various benefits offered by this approach. The Blockchain approach significantly improves a large number of implementations through the realization of effective security and accountability in the form of a distributed ledger. The authors in this research article have proposed the use of Blockchain for the purpose of achieving effective data vending. The implementation of this distributed approach significantly improves the data vending characteristics and introduces and accountability through the implementation of smart contracts.

D. Peng Elaborate on the concept of the crowd sourcing and crowd sensing implementations that have been utilized extensively in the recent years. The crowd sensing approach implements the distribution of a large number of mobile devices for the purpose of collecting and sharing information with a shared goal. Crowd sensing approach is very similar to the crowd sourcing implementation as it borrows a large number of concepts to achieve greater efficiency which can be accountable and highly reliable. Due to the large amount of interest by the researchers in the crowd sensing field there is a lack of an effective incentive mechanism that glitters towards improving the data quality of the approach. Therefore the authors in this approach have proposed a data quality driven incentive approach for designing the crowd sensing implementation.

K. Yang States that there has been number of useful and highly efficient implementations that have been achieved through effective implementation of the crowdsourcing platform. The crowdsourcing approach has been significant in realization of a large number of opportunities that have been witnessed for implementation in in mobile crowdsourcing networks which have been highly successful. They also bring considerable challenges involved with achieving effective network architecture for the realization of privacy and security in such mobile networks implementing crowdsourcing approaches. All of these opportunities and challenges have been essentially evaluated for their critical privacy and security of the data in these networks.

R. Ouyang Introduces the concept of crowd sourcing that has been the area of interest for a large number of individuals that have been utilizing this platform for the purpose of achieving unconventional implementations. The crowd sourcing approach significantly improves the procedure of achieving the required data or information by effectively aggregating contributions from an insignificant group of individuals. This can be useful as it provides cost and time efficient approach for easy implementation and realization of the end goal. Therefore the authors in this approach have proposed the utilization of quantitative crowdsourcing for the purpose of achieving truth discovery in a streaming as well as a parallel implementation.

A. Azaria Discusses the paradigm of Blockchain that have been getting increasingly popular in the recent years due to the various improvements that this distributed ledger has to offer. These approaches have been significant in the realization of effective security and reliability which have been one of the most essential and considerable requirements recently. One such implementation it is proposed in this research article for the purpose of achieving effective security and privacy for electronic medical records which can contain large amounts of personally identifiable information. Therefore to improve this approach the Blockchain platform is utilized to provide a permission management and data access approach which has led to promising results in the working prototype.

J. Huang explains that the conventional techniques for achieving effective collection affirmation as well as interconnection of various sensing devices have been centralized in nature. Having a centralized system is highly prone to various setbacks and security issues that can be visible if the central node goes through a failure. This has been significant in in realizing various privacy concerns that have been considerable and need to be eliminated effectively to achieve effective improvement. Therefore the authors in this approach have proposed an effective crowd sensing system that utilizes the Blockchain platform for the realization of affective reliability and increase in the privacy off the data being transferred through these sensing systems.

J. An Expresses that there has been an explosion in the platform of internet of things with large number of devices with varied sensing models being deployed for achieving greater understanding of our world. These approaches have been significant in achieving large amount of information that has been useful in various implementations and improving our understanding considerably. But there is always been an issue of privacy and security of the data that is being collected and shared through these devices. Therefore to improve the privacy and the reliability of these crowd sensing implementations the authors have proposed the use of grading evaluation and quality control which is reliant on a Blockchain based on two consensuses.

H. Duan Elaborates on the concept of crowd intelligence that has been highly useful for achieving improvements and solving problems through the aggregation of large number of individuals of average intellect. It has been shown that the crowd intelligence can be a real and useful implementation to achieve highly insightful and intelligent solutions to problems without the interference of intelligent individuals. But this has been difficult and challenging to achieve due to various privacy concerns which need to be

eliminated. To provide a solution to this problem the authors have proposed the use of the Blockchain framework to aggregate crowd intelligence in a robust and accurate implementation.

J. Xu States that crowd intelligence effectively infers gathers or processes large amount of information from a considerately large number of individuals to achieve highly intelligent and well informed solutions for complex problems. But the implementation of this crowdsourcing approach has not been as effective due to the lack of trust between the workers and the requesters in the system. The food to improve this reliability of the system the authors have proposed the implementation of Blockchain which can safeguard the Mobile edge computing ecosystem on a trustworthy the crowd intelligence approach. The proposed methodology has been effective in achieving the theoretical analysis with a large degree of precision.

## III. PROPOSED METHODOLOGY

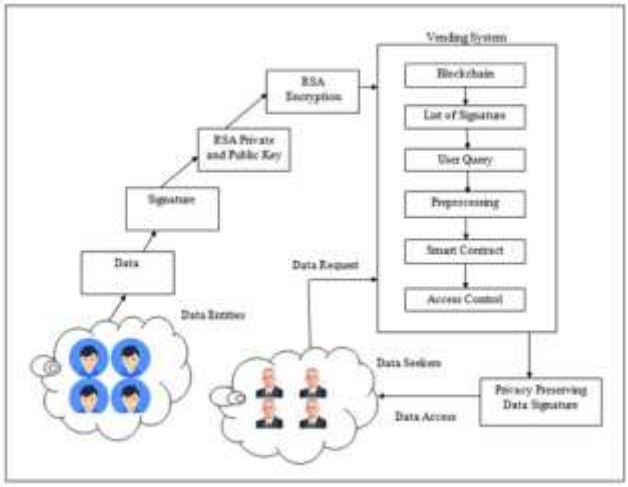


Figure 1: System Overview Diagram for Data Vending Process

The present approach for an effective data vending by the implementation of the distributed blockchain and RC encryption has been illustrated in the figure 1 above. The major steps utilised for or achieving this methodology are described below.

Step 1: Data Providers and RSA Encryption –In the first step of the presented technique the system provides an interactive user interface for the registration of both the entities the data providers and the data seekers. Once the system text input of the various user credentials the users can provide the data for selling by providing a task to the workers which can login into the system and access the tasks through their signature.

The signature password is achieved in the system through the effective utilization of the RSA encryption standard. DRS approach is an asynchronous encryption that is highly effective and extremely popular for utilization in such an implementation. In this type of encryption the public key is utilized for achieving the conversion of the plaintext into the cypher text that is the interruption process where as the private key is used for converting the cipher text into the plain text for effective decryption.

The RSA encryption considerably boosts the security of our implementation by a large margin. For achieving the encryption through the RSS standard the two pairs of keys discussed above to be effectively created for utilization. Direct approach utilizes a clever combination of cop rime and prime numbers for achieving effective generation of the public and private keys. The size of the co-prime and prime numbers is directly related to the increase in security against brute force attacks of the encryption approach. The public key can be effectively evaluated as a pair of keys and the private keys are actually generated through modulus operations as depicted in equations 1 and 2 given below.

CD= PE MOD N _____ (1)

DD= CDDMOD N _____ (2)

Where
CD - Cipher Data, P- Plain data DD-Decrypted Data

Step 2: Blockchain Formation – This is one of the most integral parts of our proposed methodology as it provides the entire effective realization of security and reliability in this approach of data vending. The data that is being uploaded by the data provider is effectively created in the form of a Blockchain for improving the reliability and security of the uploaded data. The Blockchain approach has two parts the block and the head. The head is utilized to store the hash key of the current block as well as the previous block which is effectively calculated in this approach through md5 hashing algorithm.

The data in the form of blocks is received in encrypted format from the previous step and the hash keys are calculated through the hashing algorithm. The huskies of the previous block are also added to the data of the current block and then the hash key of the next consecutive block is calculated. This effectively secures the data and forms an interconnection between the blocks through the

use of hash keys which is referred to as the Blockchain. The final key obtained when the final block of the data is stored is called as the terminal key which should be used for the purpose of authentication of the stored data on the distributed network.

The entire procedure of the Blockchain formation has been elaborated in the algorithm 1 below.

ALGORITHM 1: Blockchain formation for stored files

```
// Input: File Set FSET
// Output: Authentication Key AKEY
1:  Start
2:  Initialize Head key as HK=NULL
3:     For i = 0 to size of FSET
4:                path= FSET[i]
5                 FCONT = getFileContent(path)
6:                HK = HK +FCONT
7:                          AKEY = HK
8:        End For
9: return AKEY
```

Step 3: Signature and advertising the Data – The data that is being uploaded it is provided as an input to this step of the process. The abstract of the data is needed so that the data seekers can potentially search the data. This data is advertised by the data providers on the servers and at the same time the signature keys for the data are generated and distributed to the buyers for the purpose of description of the uploaded data.

The signature keys generation has been depicted in the algorithm 2 below.

Algorithm 2: Signature Key Generation

```
// Input :  Data String DSTR
// Output : Signature  Key SKEY
Function : signatureKeyGenerator(DSTR )
0: Start
1: SKEY =Ø
2:  HashKey HKEY=MD5 (DSTR)
3:        N=HKEY MOD  7
4:     If  N<7, then
5:     P=N+1
6:       for  i=0 to SKEY length < 7
7:            i=i+P
8:          if  i < HKEY length, then
9:          SKEY= SKEY+ HKEY [i]
10:         HKEY =rotate(HKEY )
11:         end if
12:         else
13:         i=0
14:     end For
15:     end if
16: return SKEY
17: Stop
```

Step 4: Preprocessing – The preprocessing is the most essential and highly effective step in this methodology which immensely enhances the efficiency of the entire procedure. This is a crucial aspect as the string provided to the system without preprocessing can take a large number of resources such as time and computational resources for achieving the processing. The input query provided by the buyers is effectively preprocessed through the following steps.

Special symbol Removal – The query is effectively utilized as an input in this step of the procedure to remove any special symbols and replace them with a space. The symbols such as !, . ? Are eliminated and provided to the next step of preprocessing.

Tokenization – This procedure is highly useful for the processing in the next step of the approach as it effectively divides the string into the form of tokens which can now be referred to as a well index string that can be easily addressed by the rest of the system.

Stopword Removing – The stop words are words in the English language that are utilized for providing effective flow of conversation and conjunction between a pair of sentences. These are not necessary for the implementation in our methodology and are redundant elements that are effectively eliminated.

Stemming – Most of the words in English language have the same meaning but have different forms that are achieved by adding suffixes and prefixes to the root word. In this step of the preprocessing all of these words in the query are returned to their root word through removal of the suffixes such as going will be converted to go.

Step 5: Smart Contract – the smart contract enables an effective implementation of security between the buyer off the data and the seller. Once the data that is required by the data speaker is selected by caring the system a smart contract is initiated between the owner and the user. The smart contract procedure on initiation utilizes the user profile and the user attributes to create a unique user profile key UPK. Similar key is generated for the data provider at the same time and named as the owner key OPK.

These two keys are utilized to create the smart contract key or SCK. The smart contract key is the one utilized to provide access control to the user for the data that is being shared by the owner. The procedure to achieve this smart contract key is depicted in the equation 3 given below.

SCK = UPK + OPK -------- (1)

This generation of the smart contract key is highly useful as it allows for effective achievement of securing the data which can only be decrypted by the smart contract key SCK. This effectively secures the data as the two keys needed for the generation of the smart contract key cannot be accessed by the user to misuse the data in any way. This enhances the trust between the various actors of the system and provides a trust less way for data vending in this methodology.

## IV. RESULTS AND DISCUSSION

The proposed approach for the purpose of achieving an effective data vending system has been developed through the use of Java programming language. For achieving the programming of the system the NetBeans development environment it has been used. The system is installed on a laptop consisting of Intel i5 processor assisted through to a 500 GB of hard drive and 4GB of primary memory. The database responsibilities are handled by the MySQL database approach.

The proposed technique for data vending has been effectively put under the hammer for extracting the performance of the approach through rigorous experimentation. The evaluation methodologies have been listed below.

**Character assignment for Encryption Comparison –**

The cryptographic technique utilized in this methodology implements a pair of keys for the encryption purposes. The public key is used for encryption whereas the private key is used for decryption. These are unique keys that are generated through extensive procedures elaborated in this research article previously. This process for generation needs to be evaluated by checking the performance of the approach.

The performance of the cryptographic algorithm RSA is achieved through experimentation using increasing number of characters. These characters are encrypted and decrypted and the keys generated for this process are counted for the unique number of characters. This evaluation is listed in the table below.

| Experiment No | No of Characters | No of Characters (RSA) | No ofCharacters (RCC) |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 2 | 1000 | 52 | 40 |
| 3 | 2000 | 54 | 51 |
| 4 | 3000 | 59 | 52 |
| 5 | 4000 | 57 | 55 |
| 6 | 5000 | 61 | 58 |

Table 1: Comparison of RSA and RCC for number of characters

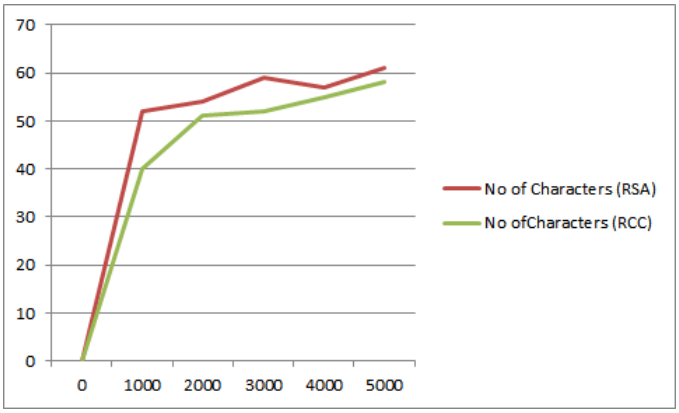


Figure 2: Plot of the number of characters utilized for encryption in RSA v/s RCC encryption techniques.

The values achieved in the table above are illustrated diagrammatically in the figure 2 below. As it is evident from the valuation a significantly large number of characters are utilized when correlated to the approach of RCC encryption depicted in [11]. This proves the improved accuracy and robust security of the encryption approach device in our system.

**Query Searching time Performance**

The performance for the search performed in our system has also been effectively evaluated and compared with the approach devised in [12]. The data provider uploads the data onto the distributed server and advertises it for anyone that is interested in the data. This data is effectively queried by the interested data Seeker through passing the relevant query to the search module of our approach. The performance for searching and retrieving the required data is evaluated and the outcomes are effectively elaborated in the table 2 given below.

| Number of Keywords in Search Query | Search time (in Seconds) |
|---|---|
| 1 | 8.55 |
| 2 | 8.76 |
| 3 | 8.86 |
| 4 | 9.08 |
| 5 | 9.23 |
| 6 | 9.33 |
| 7 | 9.45 |

Table 2: Time Evaluation for Search Query

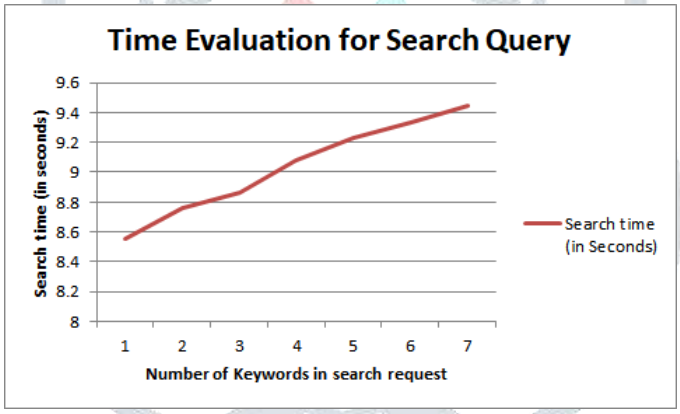


Figure 3: Graphical Representation of Time Evaluation for Search Query

The evaluation of the time elapsed for achieving the search of the query passed to our system has been GNE performed on encrypted data through Blockchain which is elaborated in [12]. The methodology proposed in [12] achieves 15.57 seconds of average search execution time whereas our approach attends an average search time for the past query of 9.03 seconds. This significant difference in time can be attributed to the trapdoor generation technique utilized in our methodology which is far superior to the one illustrated in [12].

## V. CONCLUSION AND FUTURE SCOPE

The proposed methodology for an effective and secure data vending approach is achieved through the implementation of the Blockchain platform. There are two entities in this system where one is the data vendor which provides the data and other is the data seeker which vi the data and are effectively registered into our system through an interactive user interface. The data uploaded by the data provider is effectively encrypted through RSA and converted into a Blockchain for improving the security by large margin. This data can be effectively searched through the user interface by the data Seeker and the system retrieves the relevant information in a quick manner. The data is shared with the data Seeker through evaluation of the smart contract which secures the data even further. The experimental results have been compared with conventional approaches and have had a considerable improvement in various models such as the caring time for search performance and effective implementation of the RSA encryption.

The future work can be in the direction of enhancing this methodology for the through implementation on a cloud based service for effective integration and ease of use.

## REFERENCES

**[1]** D. Dang et al, "A Crowdsourcing Worker Quality Evaluation Algorithm on MapReduce for Big Data Applications", IEEE Transactions on Parallel and Distributed Systems, 2016.

**[2]** J. Zhou et al, "Distributed Data Vending on Blockchain", IEEE International Conference on Internet of Things (iThings), 2018.

**[3]** D. Peng et al, "Data Quality Guided Incentive Mechanism Design for Crowdsensing", IEEE Transactions on Mobile Computing, 2017.

**[4]** K. Yang et al, "Security and Privacy in Mobile Crowdsourcing Networks: Challenges and Opportunities", IEEE Communications Magazine, 2015.

**[5]** R. Ouyang et al, "Parallel and Streaming Truth Discovery in Large-Scale Quantitative Crowdsourcing", IEEE Transactions on Parallel and Distributed Systems, 2016.

**[6]** A. Azaria et al, "MedRec: Using Blockchain for Medical Data Access and Permission Management", 2nd International Conference on Open and Big Data, 2016.

**[7]** J. Huang et al, "Blockchain-based Crowd-sensing System", 1st IEEE International Conference on Hot Information-Centric Networking, HotICN 2018.

**[8]** J. An et al, "Crowdsensing Quality Control and Grading Evaluation based on a Tw o-consensus Blockchain", IEEE Internet of Things Journal, 2018.

**[9]** H. Duan et al, "Aggregating Crowd Wisdom via Blockchain: A Private, Correct, and Robust Realization", IEEE International Conference on Pervasive Computing and Communications (PerCom), 2019.

**[10]** J. Xu et al, "A Blockchain-enabled Trustless Crowd-Intelligence Ecosystem on Mobile Edge Computing", IEEE Transactions on Industrial Informatics, 2019.

**[11]** Ebenezer R.H.P. Isaac, Joseph H.R. Isaac and J. Visumathi, "Reverse Circle Cipher for Personal and Network Security", 2013 International Conference on Information Communication and Embedded Systems (ICICES), 29 April 2013.

**[12]** Shan Jiang, Jiannong Cao, Julie A. McCann, Yanni Yang, Yang Liu, Xiaoqing Wang and Yuming Deng, " Privacy-preserving and Efficient Multi-keyword Search Over Encrypted Data on Blockchain", IEEE International Conference on Blockchain (Blockchain),2019.