International Journal of Research in Engineering, Science and Management
Volume 4, Issue 6, June 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Distributed Data Vending through Crowdsourcing based on the Blockchain Framework

Stephen Dias[1*], Shubham Gawade[2], Pranav Goel[3], Piyush Bhujbal[4], Balaji Bodkhe[5], Mahesh Shinde[6]
[1,2,3,4]*Student, Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, India*
[5,6]*Assistant Professor, Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, India*

***Abstract*: Data is an essential and highly useful commodity in this information age. With the high abundance of data in becoming is very difficult to achieve the relevant data for various implementations. Machine learning and artificial intelligence approaches require massive amounts of data to achieve their predictions and classification goals. With the massive amounts of data being generated, there is a lack of trust between various data providers as well as seekers which leads to an inability of these individuals to work together efficiently. Therefore, in this research article and effective crowdsourced implementation for a data vending approach has been specified. The crowdsourcing approach significantly improves the data binding approach, as well as the security of the data, which is preserved through effective encryption implementations. The presented technique utilizes encryption in the form of reverse circle cipher and entropy estimation along with the implementation of the distributed blockchain framework and decision tree. The methodology has been effectively quantified for their performance metric through the implementation of extensive experimentation which has proven the superiority of the proposed methodology.**

***Keywords*: blockchain, reverse circle cipher, data integrity, hashing, task worker, task provider, crowd intelligence, data vending.**

1. Introduction

Data is a highly valuable commodity in this world, as the age of information has dawned on human beings. Data collection and proliferation have been useful in imparting much-needed knowledge to humans. Most of this information in the early ages was facilitated and distributed through the word of mouth. This has been indicated in the various cave paintings of the Stone Age Era that have been discovered. Humans have improved upon the knowledge of their ancestors through the effective transfer of data. Every piece of information is highly useful as it can be significantly improved in upcoming years to achieve highly improbable goals. This is how a lot of technology, as well as other improvements, have been happening across the world.

The data collection approaches have been significantly improved in recent years. Earlier one of the most effective approaches to store and distribute data was through the use of books. Books are still used as one of the most useful approaches for transferring data and a lot of students and other individuals utilize books to help them gain insightful knowledge and information. Various technological advancements have improved significantly to achieve permeance of data through electronic storage options. For this purpose, various magnetic disk solid-state drives and other flash storage have been devised. These approaches have been effective in achieving improved storage densities have been getting enhanced every year.

With the development of the internet platform, there has been an exponential increase in the amount of data that is being generated worldwide. The internet platform has been a host to a large number of web applications and web pages that have been significant in achieving the various amount of services. These services include cloud service, social media, e-commerce, online banking, etc. These approaches generate massive amounts of data every single day. This data is highly useful as it contains valuable metrics and statistics that can be effectively processed to achieve insightful information regarding the various fields and their implementations. This data is highly valuable in the implementations of machine learning and artificial intelligence approaches which can be significantly useful in achieving effective processing for a variety of implementations.

Therefore, in various approaches to successfully use this data and achieve highly effective realizations. But there has been a unanimous lack of trust between various entities that have led to most of the data-sharing approaches not being viable or utilized due to these trust issues. This is compounded by the fact that most of this information contains personally identifiable information that can be highly problematic if it is leaked. This lack of trust has been highly damaging for the various

*Corresponding author: stephendias101199@gmail.com

implementations that require maximum amounts of data to achieve viable results and insightful realization. Therefore, there is a need for an effective and useful data sharing data vending approach that can be utilized for sharing such data in a trustworthy and reliable environment.

The paradigm of crowdsourcing has come to the rescue for this approach as the crowdsourcing approach can be highly useful in the collection and dissemination of data in a large-scale manner. The crowdsourcing approach utilizes a large number of individuals to perform a single and complicated task. This is useful at the large and complicated task can be effectively divided into smaller and easy parts that can be effectively utilized by the individual workers to complete the task. But the crowdsourcing approach has also been played with a lot of accountability issues that have reason between the task provider and the workers. This approach is also in the need for or improvement in the reliability to considerably enhance the trust in the various actors involved in this approach.

Therefore, this methodology achieves a trustworthy implementation of a crowdsourced data vending approach by the implementation of the distributed blockchain framework. The blockchain approach is one of the most essential and useful realizations of a distributed ledger that can effectively prevent any tampering of the transactions. This approach utilizes the data in the form of blocks that are chained together through the utilization of hash keys. These blocks of data if tampered could lead to the breakage of the chain which can be highly evident which can significantly improve the reliability and security of the implementations realizing this approach. The smart contract approach of blockchain has also been implemented to significantly e enhances the security between the provider and the worker by a large margin. The specified approach is effectively elaborated in this research article in the upcoming sections in detail.

This research paper for achieving trustless crowd intelligence system for Data Vending through blockchains is dedicated segment 2 for Literature Survey. Section 3 narrates the steps that are carried out in the process of managing crowd intelligence and Data vending in Worker and Provider paradigm. Section 4 of this research paper evaluates the obtained results with the other existing methods. Finally, section 5 concludes this paper along with the options to enhance the same in the future.

2. Literature Survey

H. Zhang [1] explains that decision making is one of the most essential aspects for human being. Decision making comes naturally to individual that is an expert in that particular domain. What various challenges in different environments are not usually considered while making these decisions. Therefore, the authors have proposed the utilization of a crowd intelligence platform for the purpose of achieving effective decision making. The decisions in this platform are achieved through the comparison between negative and positive based on the linguistic scale. The consensus measure and fuzzy numbers play an important role in achieving this effective approach for crowd intelligence.

R. Shit [2] elaborates on the various improvements being made in achieving effective and useful transportation which is sustainable in nature. For achieving and effective system for transportation systems in the future there is a need for an interactive approach that can be utilized to achieve this objective. This is due to the fact that there are large number of individuals that are connected to the transport system and their input is highly valuable in achieving the prescribed goals. Therefore, the authors propose a futuristic transport system that is intelligent in nature through the implementation of crowd intelligence. The authors convey that the implementation of crowd intelligence can significantly improve the uh parking solutions traffic protections traffic control and mobility which could lead to the transportation system in achieving a smart approach easily.

J. Zhang [3] expresses the fact that there has been a large amount of interest be bestowed on the open innovation topic. The open innovation allows for an effective and useful improvement in the research capabilities due to the rapid increase in the amount of information in the current age. This is further enhanced through the realization of the crowd intelligence approach that further bolsters the system to achieve even better and useful implementations. Therefore, for this purpose the authors have performed an extensive study of the implementation of crowd intelligence in the open innovation community for their design patterns.

Y. Yang [4] discusses the increased importance being paid to the collaborative systems that have been known to achieve greater efficiency and reliability. These systems are highly useful as it allows for an effective increase without any known problems or negative effects. The crowd sourcing approach significantly improves these approaches due to the realization of the internet platform. Crowdsourcing has been highly useful in various approaches to achieve an effective and large impact through minor contributions from a large number of individuals. Therefore, the author proposes an effective technique for collaboration of crowd for creating and designing a platform through the realization of big data analytics.

L. Rosenberg [5] introduces the concept of collective intelligence in the form of a story that is highly popular across the world. The story recounts the incident of estimating the weight of an animal by large number of individuals that that were not experts but rather just common people. It was expected that most of these individuals would not be useful for achieving the weight of the animal easily. But taking the average of about 780 estimates resulted in the weight of the animal almost perfect. This story was utilized for the purpose of depicting the wisdom of large number of individuals that have been used in this study for the effective comparison of intelligence of crowds or swarms to achieve highly useful results through crowd intelligence.

G. Willcox [6] narrates the well-known idea that a large number of people can corporate for a particular topic or interest reaching intelligent realizations and results through crowd participation. This paradigm has been closely studied in this research article for the purpose of quantifying the ability of crowd intelligence effectively. The authors indicate that the

crowd intelligence approach have been noticed in the natural world increasingly. There are large groups of animals that have been highly useful in improving the survivability as well as the intelligence of the entire group by a large margin. These groups are referred to as hive minds by the authors and a referred as closed loop system that can enhance the performance on a particular task bye by an extremely large margin to achieve effective decision making capabilities.

J. Leng [7] explains that it there has been an increase in demand of various products across the world which has been directly linked to the growth of population and industrialization. This growing demands are highly difficult to keep up as various industries have been clamoring to achieve the desired demand supply ratio. This is further complicated by the fact that there is a lack of an effective system that can analyze the approach effectively and reduce the instances of large-scale disparity in the demand and the supply. Therefore, the authors in this approach have proposed ManuChain as an effective approach for implementing bi level intelligence through a holistic optimization model on a permission blockchain distributed framework for smart manufacturing. This approach has been effective in improving the approach significantly through the realization of the blockchain framework.

Y. Kano [8] elaborates on the blockchain approach that has been significant in achieving effective and useful currency in the digital realm. This has been useful in achieving significant attraction by a large number of individuals which has been growing exponentially in the recent years. This is due to the effective temper proof nature of the blockchain platform that provides effective security to the transactions performed on Bitcoin. Therefore, the authors in this approach propose a solution to the concentration problem in the mining work to mine blockchain based currencies. The authors propose and alternative technique for mining the blockchain through the utilization of mobile computing environments in a ubiquitous manner. The authors propose the utilization of crowd intelligence in a gamification approach to achieve significant centralization of the mining task.

K. Singi [9] expresses that there has been increased Reliance on software's and other applications that are being utilized for the purpose of supporting businesses and their clients on the digital platform. This is necessary at is as it allows various functional specifications and business strategies to be applied and implemented easily through the use of various tools and external libraries. The life cycle of software development in various business derived software's is highly focused on the client. But due to the large number of different individuals and other aspects this leads to a drift between the requirement it and achievement of the various goals. For this purpose the authors propose an effective technique for the compliance and adherence along with governance in the delivery of software to the implementation of blockchain and the crowdsourced platform.

J. Huang [10] discusses the role of industrial systems and the rise in the mobile sensor approach that has been useful in achieving these industrial systems. These industrial systems have been improved by deploying statically sensors that have been highly reliable for achieving sensory data for these industrial systems that are data driven. These sensory systems are highly useful in achieving effective realization of the goals but are heavily constrained as the cost of maintenance and deployment is very high. The mobile sensing approach has been a promising development that has been useful in achieving effective scalability and mobility along with the cost reduction effectively. Therefore, the answers have proposed utilization of a mobile crowd sensing approach based on the blockchain platform to improve the security significantly in application on industrial systems with very high effectiveness.

J. Xu [11] introduces the concept of internet of things and the management of these devices that have been highly useful in various application across different paradigms and fields. The internet of things approach is highly useful as it allows effective improvement in the various current systems that have been implemented by increasing their efficiency and preciseness significantly. But due to the nation see of this approach and the different limitations imposed due to the nature of these devices the security of these approaches has been something that is left to desire. For this purpose, the researchers in this approach have been useful in outlining and effective approach for implementing h computing platform for internet of things applications that are intelligent in nature on a distributed blockchain platform. This approach significantly improves the security of the approach and the internet of things through the realization of the blockchain approach.

K. Xin [12] narrates that the crowd sourcing approaches has been highly useful in achieving various complicated tasks through small contributions from a large number of individuals in the crowd. This concept is highly useful as there isn't a need for a very intelligent worker but a large number of moderately intelligent people or workers can solve an extremely difficult and complicated task easily through the division of the task in small easily accessible goals. But the author indicates that most of these approaches for crowd intelligence have been defunct due to lack of accountability and reliability on the platform as well as lack of trust in the workers. The effort to improve this paradigm the authors have proposed a crowd sourcing approach that provides reciprocal implementation for collaborating on a certain task in a transparent manner through the implementation of the blockchain approach. The authors of quantified this approach through the use of a blockchain game called cell evolution and achieved highly satisfactory results.

3. Proposed Methodology

The presented technique for an effective and useful data vending approach through the realisation of accountability by the implementation of reward and penalty scheme has been depicted in the figure one above.

This procedure is performed in a sequence of steps that are listed in detail below.

Step 1: Registration and Activation of the System: The proposed methodology has been effectively developed for the purpose of implementation on three different machines. This approach can similarly be realized on a single machine through simulation of the other two actors. This is due to the fact that

the methodology requires three different roles one of the decentralized server second the task provider and third the worker. Administrator of the system maintains the decentralized server through the utilization of an interactive front end through the use of swings framework based on the Java programming language.

The task provider as well as the workers utilize this interactive user interface for the purpose of facilitating the registration of these entities. The workers and attach providers utilize their relevant attributes for the purpose of registration before they can use the system. The attributes considered for the registration purposes are name, mobile number, user name and password along with the email address. Once the attributes entered by the user are validated the respective registration is performed for the chosen role in addition to signature key generation.

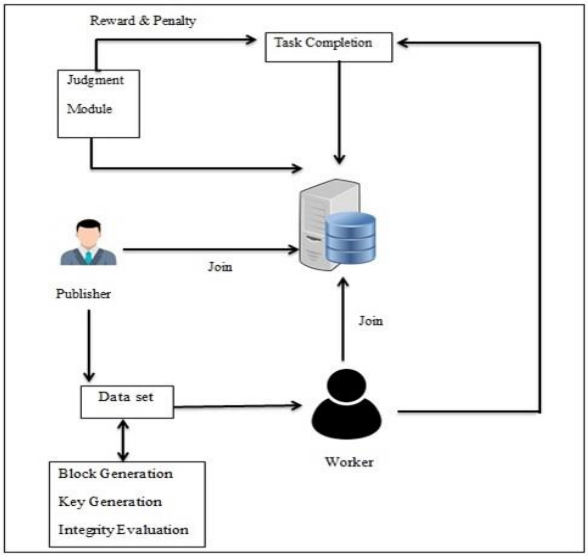

Fig. 1. System overview

For the purpose of enacting the rules the respective users can achieve authorized entry into the system through the utilization of the unique username and password provided at the time of registration. The contact related user attributes are utilized to generate hash key which is then used to select random characters for the creation of the signature key. The signature key generation has been elaborated in the algorithm 1 given below.

ALGORITHM 1: Signature Key Generation

//Input: Attribute List A_{TL}
//Output: Signature Key KEY
Function: *keyGeneration*(A_{TL})
1: Start
2: con = " ", KEY= ""
3: *for* i=0 to size of A_{TL}
4: con = con + $A_{TL[i]}$
5: *end for*
6: $MD5_{HK}$=MD5(con)
7: REM= $MD5_{HK}$ SIZE MOD 7
8: *for* j=0 to KEY Length < 7
9: j=j+(REM+1)
10: *if* (j< $MD5_{HK}$ length)
11: KEY=KEY+ $MD5_{HK\,[j]}$
12: $MD5_{HK}$ = $MD5_{HK}$ >> 1
13: *else*
14: j=0
15: *end for*
16: return KEY
17: Stop

Step 2: Uploading Task and RCC encryption: The data provider or publisher utilizes the system for the purpose of uploading the task off the data for the workers or the data requesters onto the distributed system. The crowdsourcing approach is effectively implemented through the provision of providing the task to the workers by the task provider. The workers are seeking the data and the publisher is a need for workers that are requesting the work. This allows an effective and useful realization of the crowdsourcing platform that provides benefits to both the entities that are involved. Therefore, this system there is a need for a publisher to provide the task all the data to the worker by the utilization of the decentralized server.

The task that is being published by the publisher is uploaded on to the distributed system in the format of a text file. This text file consists of a number of attributes namely the name of the task, description, price of the task, and the security deposit needed from the worker. This task in the format of a text file is successfully provided to the distributed server for uploading. The task needs to be effectively encrypted before uploading to ensure the security is maintained intact.

For the purpose of encryption, the reverse circle cipher approach is being utilized. For the purpose of uploading the text file containing the task provided to the system as an input. The input path is traversed by the system and the relevant text file is extracted from the given location. The extracted text file is effectively processed and the various attributes and the contents of the file are extracted by the system and converted into to the string format. This is useful as a system can effectively process strings and utilize it for encryption and further processing.

For the purpose of encryption, the signature key created while registration of the entities is utilized. The key is processed by the extraction of the characters in the key and converting them into their respective ASCII values. This ASCII values are then summed up together to achieve a resultant value. This value is then modded by 20 which results in the add up value. This value is highly useful as it is effectively utilized for achieving the encryption. The string is first divided into blocks through the equation 1 given below.

$$f(RSB) = \int_0^n SB(i) \Rightarrow SB'(i) \quad (1)$$

Where,
R_{SB}= Rotated Division List
n= Number of blocks
S_{Bi}=i^{th} Block
S_{Bi}'=Rotated i^{th} Division

The resultant string created through the extraction of the contents in the text file provided by the task provider is segregated into blocks. Which of this block contains 10 characters and these blocks a stored individually in the form of a list. The blocks are they subjected to rotation when the characters are rotated according to the index of the block. The number of rotations is achieved by modding the index of the block by 10.

Once the blocks are rotated, the characters in these blocks utilised to extract their ASCII values. This ASCII value of the characters is then summed up with the art of value extracted from the signature key in the previous step. This results in a new ASCII value of the resultant block which is then assigned and replaces the characters in the block resulting in efficient encryption which is depicted in equation 2 below.

$$f(ESB) = \int_0^n \sum_0^k SB(i)\ \ MOD\ Ky \Rightarrow ESBi \quad (2)$$

Where,

E_{SB}= Encrypted Block
n= Number of blocks
k= Numbers of Character in the block
S_{Bi}=i^{th} Block
K_y= Cipher Key
E_{SBi}=Encrypted Block

This resultant list is effectively utilised to generate the text file by concatenating the blocks and the resultant encrypted text file is uploaded through socket programming on to the distributed server. The entire procedure of reverse circle cypher is illustrated in the algorithm 2 below.

ALGORITHM 2: Reverse Circle Cipher

// Input: File String F_S
// Output: Cipher Data CIP_{STR}
Function reverseCircleCipher (F_S, KEY)
1: Start
2: Initialize list Block LST_{BLK} =∅, DIV_{STR}="", addupval=0
3: *for* i = 0 to size of KEY
4: addupval= addupval+ASCII ($KEY_{[i]}$)
5: *end for*
6: addupval= addupval MOD 20
7: *for* i = 0 to size of F_S
8: char ch=$F_{S\,[i]}$
9: DIV_{STR} = DIV_{STR} +ch
10: *if* (DIV_{STR} size =10), *then*
11: LST_{BLK} = LST_{BLK} + DIV_{STR}
12: DIV_{STR} =""
13: *end if*
14: *end for*
15: LST_{BLK} = LST_{BLK} + DIV_{STR}
16:
17: For i = 0 to size of LST_{BLK}
18: STR= $LST_{BLK\,[i]}$
19: STR=rotate (STR, i)
20: For j = 0 to size of STR
21: char ch= $STR_{[j]}$
22: newchar=ASCII(ch) + addupval
23: CIP_{STR} = CIP_{STR} +newchar
24: *end for*
25: *end for*
26: *return* CIP_{STR}
27: STOP

Step 3: Blockchain Creation and Integrity Maintenance: This is the most essential step of the presented technique. This type is utilized to achieve the maintenance of the integrity off the data uploaded on the decentralized server. The decentralized server is utilized for the purpose of storing the encrypted file obtained from the previous step in this approach. This step of the approach first performs a check if there are any files in the path that is used for the storage.

If the path is identified to be empty, then and the respective file is written in the destination and the contents of this file are extracted. The md5 hashing module is utilized for the purpose of performing hashing of the file content in the destination folder. Bhaskar the file contents are achieved by splitting the content and generating the respective hash key. The algorithm 1 is utilized to perform the creation of the head key through the hash key generated of the file contents. This head key is highly useful for the blockchain creation. This head key is that stored into the database table for the future integrity evaluation and verification purposes.

On the other hand, if there are already files on the digital a server and the destination is not empty the data contained in the server needs to be verified for its integrity. This is due to the fact that there should be an effective integrity maintenance every time a file is being uploaded on to the distributed server. Therefore, for the integrity valuation of the stored files the content of the files are extracted and subjected to hash key evaluation through the md5 hashing algorithm. The hash key of the files are there provided to the algorithm one for the purpose of head key creation.

The resultant head key achieved through algorithm 1 is then utilized for the purpose of concatenation with the file contents of the next file. This file contents are then and again subjected to the entire procedure of md5 hashing and had key generation. This process is iteratively repeated for all the files being uploaded on the decentralized server. The terminal ki is then achieved as the head key of the final file.

The resultant terminal key achieved from this entire procedure performed on the files uploaded on the distributed server is then compared to the terminal key previously stored in the database table. If the terminal key present in the database does not match the terminal key generated through the entire process then the server is compromised. The difference in the terminal key indicates the avalanche effect. The avalanche effect is a strong indicator of tempering being done on the data which leads to the distributed server being compromised. The new data will not be uploaded onto this distributed server as the integrity has been lapsed.

If the terminal key generated matches the terminal key stored in the database, this indicates effective integrity maintenance of the distributed server as there is no indication of an avalanche

effect. This indicates that is distributed server has not been tampered with and is safe for or future uploads as the integrity of the data stored on the server is maintained effectively. The formation of the block head and the integrity evaluation for the decentralised server can be depicted in the equation 3 and 4 mentioned below..

$$f(BH) = \int_0^n BB + PBK \Rightarrow HK \Rightarrow CK \quad (3)$$

Where,

B_H= Block Head
B_B= Block Body
P_{BK}=Previous Block Key
H_K=Hash key
C_K=Chain key
n= Number of blocks

$$f(VDI) = \int_0^n (CK = CK') \quad (4)$$

Where,

V_{DI}= Decentralized Data Integrity
n= Number of blocks
C_K=Previous Chain Key
$C_{K'}$=Current Chain Key

As the integrity of the decentralized server is confirmed the new file can be effectively uploaded on the server. This is done by extracting the contents of all the files present and calculating their head key through the hash key. This head keys are there used to create the blockchain. These head keys along with the hash key is then effectively utilized for the purpose of achieving the terminal key. The terminal key is the most essential key that determines the integrity evaluation on the resultant data being stored on the decentralized server. This terminal key is effectively stored on to the database for the feature evaluations of integrity of the stored data in the server and also while the next upload a file is being performed. The entire procedure of the blocks information is provided in the algorithm 3 below.

ALGORITHM 3: Blockchain Formation

//Input: File list $File_{LST}$
//Output: Terminal Key TRM_{KEY}
blockchainFormation($File_{LST}$)
1: Start
2: TRM_{KEY} =" "
3: *for* i=0 to size of $File_{LST}$
4: Path= $File_{LST\,[i]}$
5: F_{CONT}= getFileContent(Path)
6: $F_{CONT} = F_{CONT} + TRM_{KEY}$
7: H_K=MD5 (F_{CONT})
8: TRM_{KEY} = signatureKey (H_K)
9: *end for*
10: return TRM_{KEY}
11: Stop

Step 4: Task Accessing: For the purpose of accessing the task the worker utilizes their authorized credentials created at the time of registration in the first step of this procedure. Once the worker has been authenticated the system provides a search mechanism for the respective task uploaded on the system. The search procedure is performed by passing the relevant query into the search box regarding the type of data being required. The input query is converted into the string format and effectively preprocessed before being subjected to the search. The preprocessing is achieved effectively by the steps given below.

Special Symbol Removal: The English language provides structure through the utilisation of special symbols. This special symbols are not useful for our search procedure and can be eliminated without any problems. Therefore, the special symbols from the input query are eliminated in the first step of this preprocessing.

Tokenization: This is one of the most essential steps of the preprocessing approach and has a tangible impact on the execution of the system. This is due to the fact that tokenization converts the string into a well index string through segregation of the string based on the words.

Stopword Removal: In the English language stop words are the words that are used to provide a flow to the spoken language as well as an effective conjunction to different sentences. This are highly aesthetic in nature and do not change the meaning once eliminated from the input query.

The stop words can be oee discussed by example such as the sentence 'going to sleep' when subjected to this step of the preprocessing results in the string 'going sleep'. In this input string the word 'to' is the stop word which does not change the meaning of the sentence when removed.

Stemming: Most other words in the English language are derived from their parents words by the addition of various suffixes. This is done to provide distinction between the tenses of the words being used. The stemming approach converts the words into their respective parent words to achieve effective improvement in the processing as the word has become smaller.

For example, 'going' can be stemmed to 'go' by effectively replacing the 'ing' and reducing the footprint of the word. This procedure effectively reduces the processing time while maintaining the semantic of the word intact.

The preprocessed query is now provided to the system that performs a linear search for the extraction of the relevant task according to the input query.

*Step 5: Reward, Penalty assignment through Entropy and Decision Tree: A*fter the successful upload of the task on the system through the task provider this task can now be accessed by the worker through the search mechanism depicted previously. The worker provides the respective credentials for the purpose of authentication while logging into the system. Once these credentials are validated the worker can access the task provided by the task publisher. before the effective allotment of the task a smart contract between the worker and the task provider is initiated.

Through the utilization of this smart contract the encrypted task provided by the task provider can be effectively decrypted at the decentralized server and then again encrypted through the utilization of a combination of signature keys of the worker and

the task provider and subsequently given to the worker.

This ensures that only the respective worker can access the task and decrypt it effectively. This makes sure that the task shared by the task provider cannot be visible to anyone else other than the specified worker due to the smart contract. If the worker tries sharing this information with another person, the data cannot be decrypted effectively as it is encrypted through the utilization of a combination of keys of which only one key is available to the worker. If the worker tries to do any malpractice this can be effectively detected by the task provider easily.

If any foul play is detected the paradigm of reward and penalty is initiated which can affect the rank of the worker. The wreck of the worker is evaluated through effective implementation of entropy estimation. The Shannon information gain is utilised for the purpose of entropy estimation through the equation 5 below.

$$\mathrm{IG} = -\frac{P}{T}\log\frac{P}{T} \; -\frac{N}{T}\log\frac{N}{T} \tag{5}$$

Where,

P= Number of Positive Likes

T= Total number of likes/dislikes.

N= T-P

IG = Information Gain of the user

The resulting information gain values are useful for the updation of the rank of the worker based on their practices on the system. This information given values are provided for effective classification through the utilization of decision tree. The decision tree approach implements if-then rules to locate the deserved reward or penalty to the worker which significantly improves the reliability and accountability on this decentralized trustless data vending system.

4. Result and Discussions

The proposed methodology for an effective data vending approach through the utilization of the blockchain distributed framework has been achieved in Java programming language through the utilisation of the NetBeans integrated development environment. The laptop used for the development of this approach is powered by an Intel i5 CPU assisted by 500 GB of storage and 4GB of Ram equipped with windows operating system. The masculine database server is utilised for the purpose of fulfilling the database requirement.

The proposed methodology has been evaluated extensively for its performance on a large variety of performance metrics. The procedure for the evaluation has been depicted below

A. Encryption and Decryption Time performance

The proposed methodology utilizes encryption and decryption procedures to achieve effective security of the data being uploaded onto the decentralized server. The performance of this approach needs to be quantified which is achieved through the procedure given below. The time elapsed for the encryption and decryption procedure based on the wearing number of characters is tabulated in the table 1 below.

Table 1
Encryption and Decryption time performance

Number of Characters	Encryption Time in Milliseconds	Decryption time in Milliseconds
16	3	3
1809	14	16
2803	32	31
3001	45	53
5113	51	57
5989	63	61
6432	66	62
8246	76	77
9120	81	79
9981	95	97

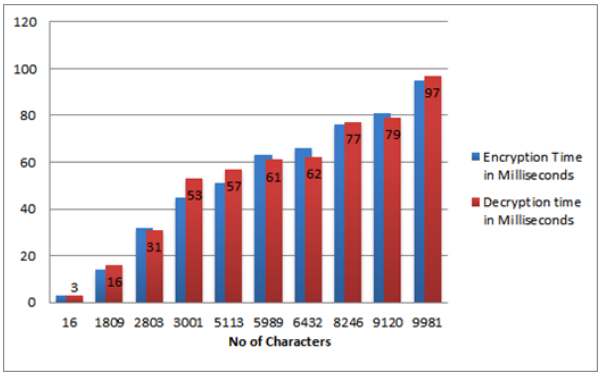

Fig. 2. Encryption and Decryption Time

The tabulated values effectively extracted for the purpose of graphical representation in the bar graph provided in figure 2 above. As it is evident the time required for the encryption and decryption process is not directly proportional to the input characters. This is due to the fact that the cryptographic process utilised for this approach that is the reverse circle cypher is is accurately developed and implemented. This is the reason why the performance of this approach is highly effective as depicted by the performance results.

B. Task Searching in Blockchain Data

For this performance metric the search performance of the presented technique has been quantified. The worker utilizes the interactive user interface for search implementation after successful authentication of the login credentials. The search is utilized for the effective extraction of the relevant data by passing the respective query to the system. The execution performance of the search query in the presented technique is evaluated by the extraction of the time elapsed for the execution of the search mechanism. A collection of queries with varying number of keywords is provided for the purpose of evaluating the search performance. The time taken for the execution of each of the search queries is tabulated in the table 2.

The figure 3 above depicts the tabulated values in the table 3 in the form of a line graph for easier representation. The time taken for the search mechanism to execute the search is effectively correlated with the approach depicted in [13]. The methodology in [13] performs multi keyword search over encrypted data through the utilization of a trap door approach.

This results in the average search time of 15.57 seconds for the specified approach in [13]. On the other hand, our methodology acquires an average of 9.78 seconds for the purpose of achieving the execution of the search query provided as input by the user. This proves the improved performance of search in our methodology in comparison to the approach depicted in [13].

Table 2
Time Evaluation for Search Query

Number of Keywords in Search Query	Search Time (in seconds)
1	9.35
2	9.56
3	9.67
4	9.78
5	9.89
6	10.09
7	10.13

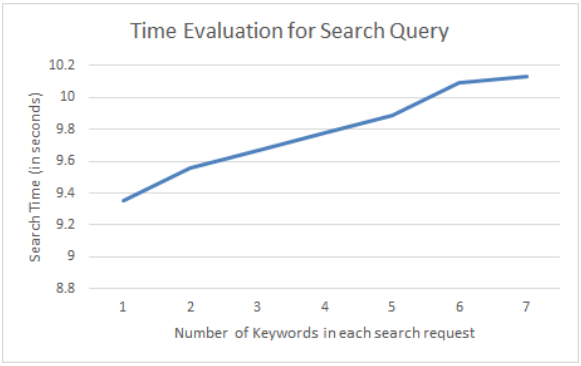

Fig. 3. Graphical Representation of Time Evaluation for Search Query

5. Conclusion and Future Scope

The methodology for an effective approach for data vending through a crowdsourcing implementation on a distributed blockchain platform has been effectively outlined in this research article. In this approach, there are three different actors first one is the decentralized server along with the task provider and the worker. The two entities are registered on the server through the effective utilization of their various attributes. Once these entities are registered the task provider upload the various task all the data on the system. The system effectively increases data using reverse cycle cipher and then forms a blockchain out of the data before uploading it onto the decentralized server. Once the data is uploaded it can be then queried by the worker according to the data that is required by the worker. When the worker of the data requester finds the requisite data through the search mechanism in this approach, then the worker requests the data from the publisher, and a smart contract is initiated which is used to encrypt the data and provided to the worker. The performance of this approach has been effectively realized through the implementation of encryption and decryption time performance and car searching performance which is compared with the conventional searching approach on encrypted data on a blockchain. The comparison has indicated that the presented methodology is superior to the conventional approaches.

For future directions of this research, this approach can be realized on a cloud platform for real-time implementation.

References

[1] H. Zhang, Q. Zheng, Z. Wang, Y. Chen, Y. Qu and T. Liu, "Crowd Intelligence for Decision Making Based on Positive and Negative Comparing with Linguistic Scale," 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2018, pp. 1-8.

[2] R. Shit, "Crowd intelligence for sustainable futuristic intelligent transportation system: a review", IET Intelligent Transport Systems, 2020.

[3] J. Zhang, L. Wang, L. Shi, W. An and W. Wei, "Study on crowd intelligence design pattern of the open innovation community," 2018 IEEE International Conference on Applied System Invention (ICASI), 2018, pp. 774-777.

[4] Y. Yang, J. Wang, W. Huang, G. Li and L. Huang, "Crowd Intelligence-Based Interactive Creation Platform," 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2019, pp. 275-280.

[5] L. Rosenberg, D. Baltaxe and N. Pescetelli, "Crowds vs swarms, a comparison of intelligence," 2016 Swarm/Human Blended Intelligence Workshop (SHBI), 2016, pp. 1-4.

[6] G. Willcox and L. Rosenberg, "Short Paper: Swarm Intelligence Amplifies the IQ of Collaborating Teams," 2019 Second International Conference on Artificial Intelligence for Industries (AI4I), 2019, pp. 111-114.

[7] J. Leng et al., "ManuChain: Combining Permissioned Blockchain with a Holistic Optimization Model as Bi-Level Intelligence for Smart Manufacturing," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 1, pp. 182-192, Jan. 2020.

[8] Y. Kano and T. Nakajima, "An alternative approach to blockchain mining work for making blockchain technologies fit to ubiquitous and mobile computing environments," 2017 Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU), 2017, pp. 1-4.

[9] K. Singi, V. Kaulgud, R. P. J. C. Bose and S. Podder, "CAG: Compliance Adherence and Governance in Software Delivery Using Blockchain," 2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB), 2019, pp. 32-39.

[10] J. Huang et al., "Blockchain-Based Mobile Crowd Sensing in Industrial Systems," in IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6553-6563, Oct. 2020.

[11] J. Xu, S. Wang, A. Zhou and F. Yang, "Edgence: A blockchain-enabled edge-computing platform for intelligent IoT-based dApps," in China Communications, vol. 17, no. 4, pp. 78-87, April 2020.

[12] K. Xin, S. Zhang, X. Wu and W. Cai, "Reciprocal Crowdsourcing: Building Cooperative Game Worlds on Blockchain," 2020 IEEE International Conference on Consumer Electronics (ICCE), 2020, pp. 1-6.

[13] S. Jiang et al., "Privacy-Preserving and Efficient Multi-Keyword Search over Encrypted Data on Blockchain," 2019 IEEE International Conference on Blockchain (Blockchain), 2019, pp. 405-410.