Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777488

A SECURE AND SCALABLE BLOCKCHAIN USING HYBRID CONSENSUS ALGORITHM WITH SMART CONTRACT FOR SUPPLY SYSTEM

V.Sarala Devi[1],Dr.S.Radha Rammohan[2]

[1]Research Scholar,Department of Computer Science and Engineering,Dr.M.G.R Educational and Research Institute, Chennai

[2]Professor,Department of Computer Applications,Dr.M.G.R Educational and Research Institute,Chennai

Abstract

Blockchain is a decentralized digital database that records transactions across a large number of computers. Block chain could be used in a variety of industries, including finance, medicine, real estate, transport, agriculture, and supply chain management. Because blockchain enables dispersed peer-to-peer connections, it ensures safety and confidentiality. Big data offers new prospects, increased value, and operational efficiencies in traditional supply chain strategies, according to this research. Hence, this paper presents a novel framework for blockchain-based Supply system with an aim to improve the scalability with secured data storage. Initially, the data set is preprocessed using normalization method and feature extraction is done by Principal Component Analysis. To improve the storage's security and node's scheduling, the proposed Precedence Partition Scheduling Algorithm is employed. We also introduce Enhanced Particle Swarm Optimization Algorithm to improve the scalability of the network. To check the nodes, we employ Delegated Proof of Stake consensus protocol. In blockchain, to change the input entity to an output entity with defined length, the message digest-5 hash approach is employed. The proposed method is analyzed in terms of performance metrics like execution time, energy consumption, CPU usage, response time, security, throughput and efficiency, and is related with the current methodologies. The presented approach was established to be scalable for supply system when compared to the conventional approaches.

Keywords: Blockchain technology, supply system ,scalability, Principal Component Analysis, Precedence Partition Scheduling Algorithm, Enhanced Particle Swarm Optimization Algorithm, Message Digest-5 Hash, Delegated Proof of Stake consensus protocol

I. INTRODUCTION

Because of unceasing discoveries throughout edge information & communication technology (ICT), nearly all company structures have been enduring tremendous changes throughout the internet era [1]. The blockchain has been a well and extremely innovative platform that has been helping to transform established company structures and create opportunities over the whole supply chain. The term 'blockchain' was typically used to describe a fully decentralized technology enabling cryptographically collecting and preserving a reliable, unchallengeable, linear event log of agreements among connected participants. Throughout the crypto currency sector, blockchain developed as just a system for processing payments. Furthermore, while blockchain technologies have established themselves as authority inside the finance market, but have just lately expanded into other fields [2]. Blockchain would be seen as a difficult obstacle as well as a potential framework. For example, blockchain could promote transparency, responsibility, and trustworthiness, as well as safety, effectiveness, and cost savings.

Because of the increasing range of firms, greater difficult transport & delivery, and tailored customer needs and desires, supply chains were getting increasingly complicated [3]. These difficulties are especially apparent in food supply chains that have socio-economical, administrative, hygienic, and environmental concerns that necessitate the use of cognitive-based interventions. That one is trustworthiness, which ensures material efficient and effective delivery, resulting in win-win scenarios in both the socio-cultural and environmental sectors.

Big data can be described as data resources with a massive density (large-scale), high velocity (movable), and high diversity (e.g., quantitative, textual, visual, and so on) that require cost-effective, creative kinds of data analysis enabling improved insight & strategic planning [4]. Because of the enhanced capacity to both acquire enormous volumes of data & apply highly effective statistical tools to large sets of data, big data analytics have lately come to significant attention. Businesses' current capacity to acquire large amounts of (various) information as well as use effective methodological approaches to that data allows them to make highly complicated conclusions that were previously based (mainly or exclusively) upon human judgment as well as intuition.

Furthermore, blockchain has been seen as a remedy to Supply system traceable issues, as well as for fostering closer and so more trusted connections not only among companies as well as respective providers, but throughout complete SCM. [5] From just one side, a blockchain-enabled smart contract (a program which may initiate a transfer) seems to have the ability to improve SCM efficacy while maintaining a distributed process. Blockchain could be integrated with certain other cutting-edge innovations (big data, the internet-of-thing (IoT), including cyber-physical platforms, amongst many others) may create negative influences in a variety of industries.

Throughout the supply chain domain, the phrases traceability & transparency have also been used simultaneously. Those phrases were related, yet they have various meanings [6]. The phrase transparency refers to the supply chain's entire visibility. The scope to which everyone of a supply chain's relevant parties get a mutual understanding of, and connect directly to, the brand data that they ask, without damage, interference, postponement, or disruption. The capacity to acquire granular data about everything that endures inside a supply chain can be referred to like traceability.

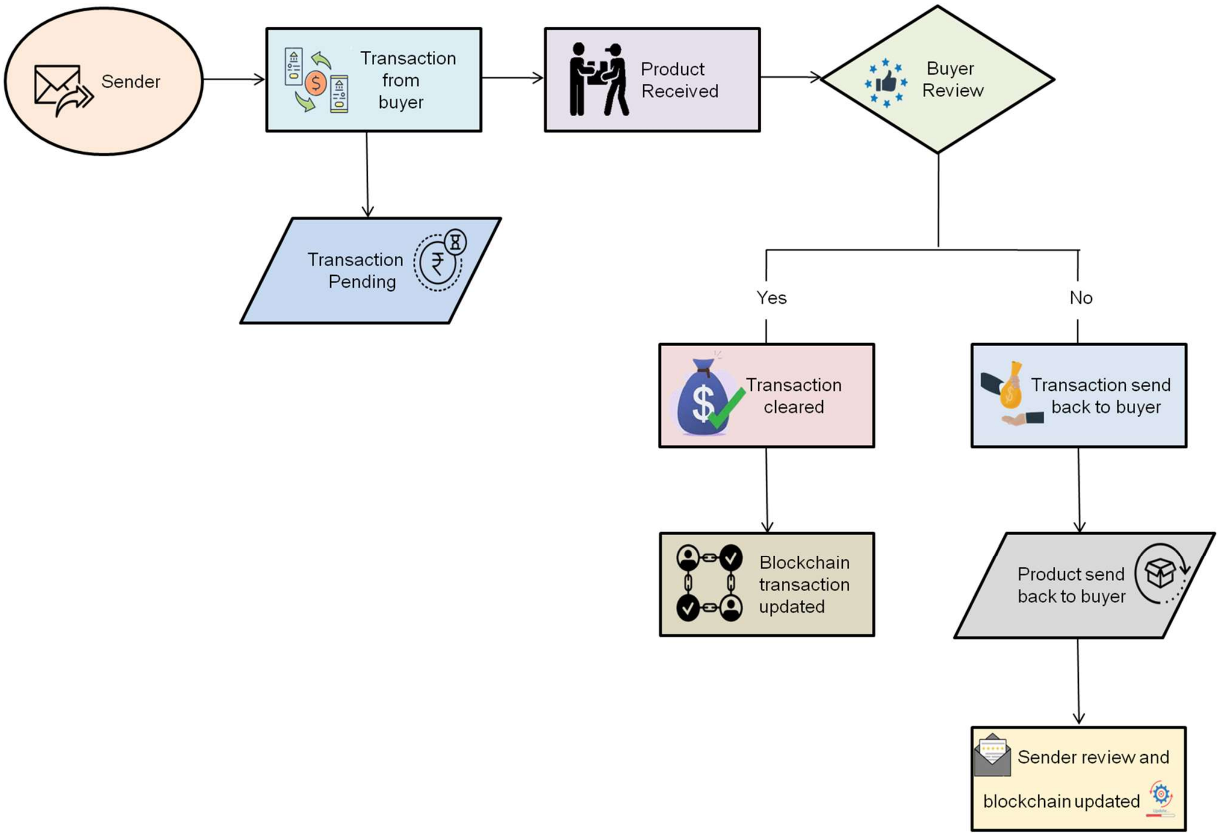
Sender
Transaction from buyer
Product Received
Buyer Review
Transaction Pending
Yes
No
Transaction cleared
Transaction send back to buyer
Blockchain transaction updated
Product send back to buyer
Sender review and blockchain updated

Figure 1: Generalized design of node to node transaction in blockchain

It could be regarding a specific asset, method, or maybe one of the supply chain's participants, including a retailer or wholesaler. Traceability could be identified on the basis of what, why, where, how, and when characteristics of the underlying asset across a supply chain in a more precise form. The phrases traceability with tracking & tracing has been closely related by investigators. A product's tracking begins at its beginning & attains its destination. The destination is often traced back to the start. History transparency has been one of 3 kinds of transparency that could be accomplished via tracking and tracing. The two sorts of openness include activities transparency & strategic transparency. Traceability, in an essence, allows for transparency via tracing as well as tracking. Figure 1depicts the generalized design of SCM in blockchain technology.

That's why this report concentrated on big data and how that can improve the supply chain operations. So, to boost scalability with secured data storage, we offer a unique approach for blockchain-based big data SCM. The data source would be first standardized, and afterwards PCA extracts features. The recommended PPSA gained the storage and node's scheduling. We further implement EPSOA for boost scalability. To define the length, MD-5 hash approach is utilized in the blockchain. We use DPoS for verify the nodes. The additional detail of our research can be structured as: topic II-literature survey with problem statement; topic III-proposed work; topic IV-result and discussion; topic V-conclusion.

II. LITERATURE SURVEY

This section explores the existing studies that prevailed in supply chain management of big data using blockchain. Also, it presents recent literature works for overcome scalability issues with effective safety in blockchain networks. [7] They developed Secure and Effective SCM design. To the authors' knowledge, it would be the first solution to use blockchain for strengthens the safety of data, logistic, and capital flows inside a supply chain system. Owing to space constraints, just the arrangement involving supply and production was already described throughout this study. [8] This article uses blockchain technology throughout the food business to measure and confirm the end product's quality efficiently. This aids offer transparency from farm to fork. This data also was important to food security organizations. This also encourages healthy competition among businesses that improve service quality. [9] They used blockchain to process supply chain data. They leveraged blockchain offline channels for scale. Assuring the safety of item serialization has been shown. This technology may reduce the availability of defective products. We'll look into using serialization to automate product recalls mostly in long term. [10] Among the most important difficulties to overcome when implementing Blockchain throughout the supply chain was including all parties. Furthermore, exchanging information across the Blockchain may cause inertia towards implementing the solution. A proper deployment of Blockchain throughout the supply chain must therefore begin with a thorough understanding of the various stakeholders' goals and requirements, in order to develop a business model that can demonstrate the solution's financial as well as client fulfillment benefits. [11] Utilizing an agent to obtain the correct response for such blockchain network (depending upon the trustworthiness employing recent consensus record) has been introduced to handle the privacy concerns created by sharding implementation. By adjusting the block size, duration, and shard count, the Deep-Q-Network agent enhances throughput. [12] Blockchain offers unique ways to solve the limits of distributed ledger, such as two-step block building in hybrid approach, effective ledger storing, and data security. Beyond from simple cargo tracking, blockchain can help with bill of lading, global commerce legality, and customs checks. [13] The distributed ledger technology offers huge ability to change the world. Nevertheless, basic study is necessary in many areas to achieve general adoption as reliable, efficient, and widespread technologies. They outlined a comprehensive framework in blockchain research method. They first defined Decentralization, Consistency, and Scalability (DCS) then showed how blockchain systems can manage these three aspects. [14] Smart logistic stability can be achieved using blockchain like a secure distributed ledger. Quantum approximation optimization & blockchain for safe heuristic optimization for smart shipping are suggested in this paper. [15] This study presents a supply chain finance authority relies on the blockchain network for highly secure storage and access management. The suggested system ensures secure preservation, fast access, and permission tracing of privacy-sensitive information whilst accommodating evolving SCF logic as well as protection needs. [16] They present a blockchain-based SCM approach which exploits smart contracts to trace as well as monitor commodities throughout the process of production. The method projects product compositions as tokens into the blockchain, providing detailed provenance details.

Problem statement

Supply Chain Formation (SCF) is the process of determining the participants in a supply chain, who will exchange what with whom, and the terms of the exchanges. Decentralized SCF appears as a highly intricate task because agents only possess local information and have limited knowledge about the capabilities of other agents. Moreover, the time consumed in transportation and supply chains from the resource acquisition to the client contributes to the business profit. Blockchain technology offers an intelligent amalgamation of distributed ledger, Peer-to-Peer, cryptography, and smart contracts to enable trustworthy applications without any third parties. The concern about significant changes in the business environment has spurred an interest in designing scalable and robust supply chains. Current approaches are not consistent because of minimum scalability throughout the SCM with blockchain. That's why this research emphasis upon big data and how it can improve existing supply chain procedures.

III. PROPOSED WORK

This article presents a unique approach for blockchain-based big data SCM. The data source would be first standardized, and afterwards PCA extracts features. The recommended PPSA gained the storage and node's scheduling. We further implement EPSOA for boost scalability. The MD-5 hash approach is applied to transform the input value into compacted numeric value. We use DPoS for verify the nodes. Figure 2 depicts the framework of our proposed work.

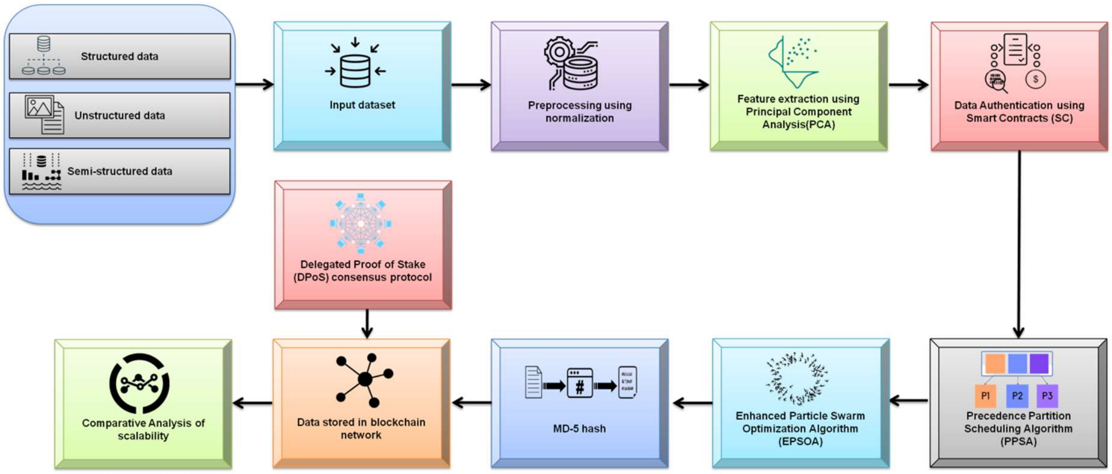
Structured data
Unstructured data
Semi-structured data
Input dataset
Preprocessing using normalization
Feature extraction using Principal Component Analysis(PCA)
Data Authentication using Smart Contracts (SC)
Delegated Proof of Stake (DPoS) consensus protocol
Comparative Analysis of scalability
Data stored in blockchain network
MD-5 hash
Enhanced Particle Swarm Optimization Algorithm (EPSOA)
P1
P2
P3
Precedence Partition Scheduling Algorithm (PPSA)

Figure 2: Framework of proposed research

A. Dataset

During 2016, we gathered information on Yangtze River delta enterprises to investigate the assumptions [17]. We conducted the survey with a local governmental organization to obtain a sample. For the administration, the organization looks at industry growth & digitalization. With such an organization's help, we got a sample comprising 1200 domestic companies. This organization's representatives gave us contact details for such companies. Firstly, authorities observed our criteria and sent out a formal request for companies to participation in the study. Then we sent out an online survey to senior executives. While we merely notified the organization and supervised the entire gathering of data, the organization's history may have caused biases toward queries of political links. For relieve respondents' worries & any biases, the online questionnaire's letter stressed the study's academic requirement.

Table-1: Structured dataset with description of samples (N=176)

Features	Samples	Proportion (%)
Company		
Metal, mechanical, and engineering	73	41.5
Pharmaceuticals	13	7.4
Petrochemicals & chemicals	24	13.6
Electricals & electronics	16	9.1
Automobiles	10	5.7
Food, alcohol, and beverages	20	11.4
Additional types of production (e.g., timber and furnishings, publication, and printings, cotton and apparels)	20	11.4
Worker		
1–99	54	30.7
100–299	79	44.9
300–499	30	17.0
500–999	13	7.4
Company's life		
Less than 5 years	23	13.1
5 to 9 years	66	37.5
10 to 24 years	84	47.7
Greater than 25 years	3	1.7

Because our research concentrated on the effect of administrative relationships in SCM, participants had to be leadership participants with specialized SCM expertise. We chose top 3 administrators (executive, production, and marketing) from every company as selected participants. The investigation used three different questions. Executive managers had to complete the questionnaires about their managerial relationships and corporate performance. For regular activities and the acceptance of Government officials & company associates, executive managers of China are suitable responders. Production managers completed the supplier integration assessment and they're in charge of supplier collaboration. Marketing managers responded to the survey on customer engagement and market instability since they are responsible for such concerns. Table 1 presents the structured dataset.

Unstructured datasets are collected in the form of documents from the Yangtze River delta enterprises. But semi-structured datasets were collected in tree-graphical formation as shown in figure 3. Here, some details are presented like name, phone number, e-mail id, etc.

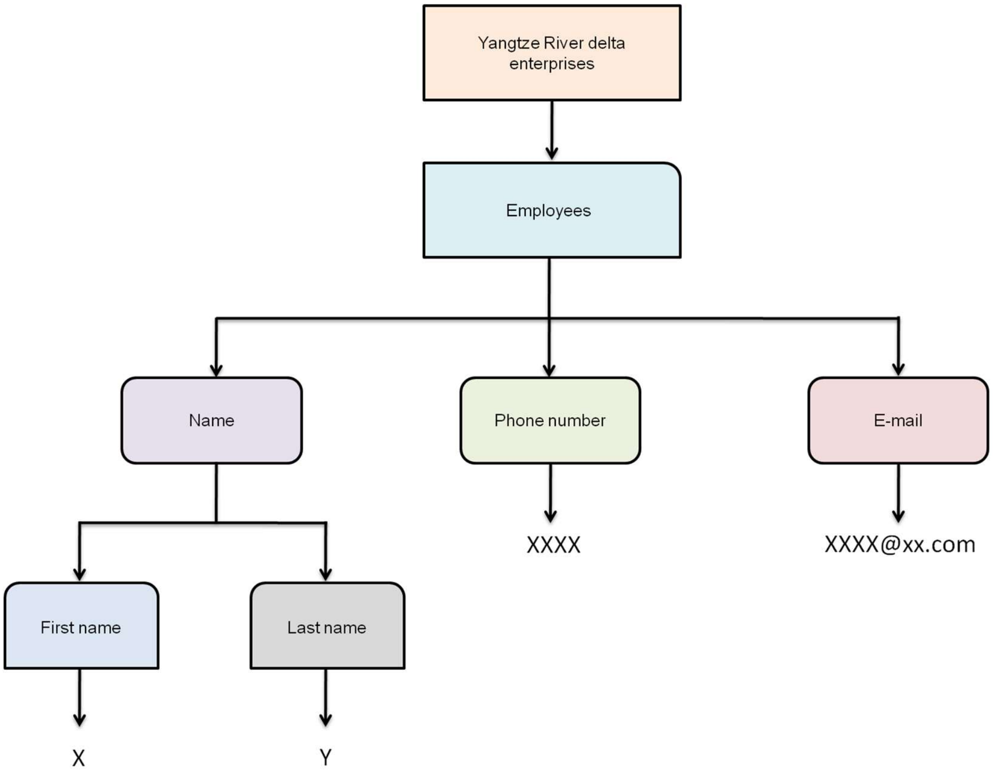

Figure 3: Formation of semi-structured dataset

B. Pre-processing using normalization

The target data should be chosen from the raw collection of financial data to boost performance. After obtaining the target data, preparation is required to create it usable. The following phase uses the fully prepared information to analyze it and generate information or results using data mining methods. Data transformation is used to change the pattern and the characteristic type. For instance, if the financial information is in numeric form yet the database needs float. Data pre-processing is an essential process of data extraction and processing.

The dataset (i.e., structured/semi-structured/unstructured) received is unfiltered and will include a forged datagram and insufficient information. It's that has been purified and normalized to delete repeated and redundant noises, along with dataset that is inadequate. Because this dataset has several features, image retrieval methods are needed to sort out the ones which aren't significant. The dataset could be normalized during the pre-processing stage. Equation (1) defines the c-count in mathematical form as,

$$\mathrm{C} = [(\mathrm{M} - \beta\ )/\tau]$$

(1)

Here, β express the mean of the information and τ hints the standard deviation. And C is represented as,

$$\mathrm{C} = \frac{\mathrm{M} - \bar{\mathrm{M}}}{\mathrm{S}}$$

(2)

Here $\bar{M}$ point out the mean of the specimen, and S points out the standard deviation of the specimens.

The random specimen looks like this:

$$C_k = \delta_0 + \delta_1 M_r + \rho_r$$
(3)

The defects that are depending on τ^2 are represented by r.

Ensuring that, as seen below, the defects should not depend on one another.

$$t_m \sim \sqrt{U}\frac{t}{\sqrt{t^2+u-1}}$$
(4)

Here, t implies the random parameter.

After that, the standard deviation is used to standardize the variable's moves. The momentary scale deviation is calculated using the formula (5).

$$\mathrm{MMS} = \frac{\mu^{\mathrm{mms}}}{\theta^{\mathrm{mms}}}$$
(5)

Here, momentary scale is denoted by mms.

$$\mu^{\mathrm{mms}} = \mathrm{Ex}(\mathrm{M} - \beta)\text{^}\mathrm{MMS}$$
(6)

Here, M stands for random variable, and Ex stands for predicted values.

$$\theta^{\mathrm{mms}} = \left(\sqrt{Ex(M - \beta)\text{^}MMS}\right)\text{^}2$$
(7)

$$t_u = \frac{mms}{\bar{M}}$$
(8)

The coefficient of variance is denoted as t_u.

The characteristic scaling procedure will be stopped by setting all of the parameters to 0 or 1. The unison-based normalizing approach is the name for this procedure. The normalized formula would look like this:

$$M' = \frac{(t-t_{min})}{(t_{max}-t_{min})}$$
(9)

The data can be kept after it has been normalized, and the length and irregularity of the data could be preserved. This phase's purpose is to minimize or wipe out data delays. The normalized dataset can then be used as feed-in subsequent steps.

C. Feature extraction using principal component analysis (PCA)

Extraction of features is indeed a step in the dimensional mitigation procedure that divides and reduces a wide range of original information into smaller groupings. As a result, processing would be simpler. To put it simply, PCA employs covariance matrix eigenvectors with the greatest eigenvalues effectively project the data into a new sub-domain with the same or fewer dimensions than the original one.

The main requirements for extracting PCA characteristics are as follows:

Stage-1: Compute the average of the images throughout the input (pre-processed) dataset.

Stage-2: To get the mean-shifted pictures, minus the average from the input dataset.

Stage-3: With diminished dimensional approach, estimate the eigenvectors & eigenvalues out of a covariance matrix.

Stage-4: In highest to the lowest, arrange the eigenvectors with its associated eigenvalues.

Stage-5: Just the eigenvectors with highest eigenvalues should be preserved (i.e. the principal components).

Stage-6: Through applying the preserved eigenvectors, initiate the mean-shifted pictures into the eigenspace.

PCA tries to extract the pre-processed data's characteristics by identifying a few orthogonal linear combinations of actual parameters that have the most variation. There are several Principal Components (PCs) even though there were initial parameters. For several data, the first few PCs illustrate the majority of the variation, allowing the remainder to be ignored with little data loss. PCA has been used to diminish the dimension of the enormous dataset while keeping as much data in the actual dataset as viable.

The mathematical relations to conduct a PCA on the data are as follows:

Capture the picture data: Presume $a_1, a_2, \ldots, a_K$ is depicted as $(M \times 1)$ vectors.

Determine the mean of the vector:

$$\bar{a} = \frac{1}{K}\sum_{i=1}^{K} a_i$$

(10)

Subtract the mean:

$$b_i = a_i - \bar{a}$$

(11)

Determine the co-variance matrix: $A=b_1, b_2,.., b_K$ from

$$V = \frac{1}{K}\sum_{m=1}^{K} b_m b_m^T = AA$$

(12)

Calculate the co-variance matrix's eigenvalues and eigenvectors

$$V = \lambda_1 > \lambda_2 > \cdots > \lambda_M \qquad \text{(Eigenvalues)}$$

(13)

$$V = c_1, c_2 \ldots c_m \qquad \text{(Eigenvectors)}$$

(14)

The pre-processed data set's principal component is the eigenvector with maximum eigenvalue. The largest eigenvalue is used to create the feature vector.

D. Data authentication using Smart Contracts (SCs)

The available SCs: Access Control Contract (AC-contract), Registered contract (R-contract), and Judging Contract (J-contract). The R-contract has been used to register participants (subjects & objects) inside the network, while the AC-contract maintains the system's total identity management. This also creates a registry table that keeps track of user details. The register table additionally manages user verification. The J-contract also influences the attitude of the subjects. It determines whether the person has engaged in any misconduct. It is termed misbehavior whenever a subject submits some many queries or rejects the created request. The J-contract imposes a punishment on the person when the misbehavior is committed. As a result, the subject's behavior determines its trustworthiness. Approval levels are tested if the person has not engaged in any misbehavior. The subject's query is then delivered to the associated object via AC-contract for obtain the necessary response. Also included is a full discussion of the AC-contract, R-contract, and J-contract.

(i) AC-contract

It's the primary smart contract throughout charge of managing accessibility. The subject transmits a message to the system when it needs a response from the object. The authorization for such subject is then maintained by AC-contract. This also improves the system's effectiveness. For every query, numerous AC-contracts are launched by objects throughout the benchmark system. The client, rather than the system, is in charge of authentication mechanism inside this system. The object deploys an AC-contract in reply to every query produced by the subject. As a response, the system's functionality and expense have skyrocketed. Furthermore, the system's computation time was impacted. Authorization between clients is governed by a unique AC-contract inside the data sharing and authentication mechanism. The AC-contract was activated whenever a subject submits a customer request to the system. Further smart contracts were run to allow customers to register and be approved. Following the customer's registration, AC-contract sends the subject's query to the object, which is checked for the appropriate authorization level.

(ii) R-contract

Clients who have been supposed to use the facilities must be genuine. Clients are authenticated via R-contracts, which are created by registering individuals inside the system. The registry table was maintained by R-contract during registrations. The table contains every one of the clients' data. R-contract generates the registry table. The registry table stores the important consumer data: subject, object, application, and duration. The registry table also keeps track of the subjects' confirmation and identification.

(iii) J-contract

The J-contract implements a judgment procedure that evaluates client behavior throughout the system. The J-contract verifies a subject's conduct whenever it submits a request inside the system. The subject will be the one who engages in the wrongdoing. It is deemed misbehaving whenever the subject submits frequent and excessive requests for services. Misbehavior also was referred as when the subject dismisses their produced request. Then, for just a subject that misbehaves, a proportional consequence was computed. The condition of the subject gets switched off for a set period of time as a punishment. If a subject's mode seems to be off, it is unable to give commands to the system. If the subject, on either side, has still not misbehaved, the authorization levels get verified. Subjects' requests are accepted or declined based on their behavior and authorization levels. Afterwards when, the J-contract sends out alert signals.

Access is allowed if the person has not engaged in any misbehavior. J-contract sends an entry permission alert notification. The J-contract, on either hand, generates various warning signals like authorization rejected or connection restricted if indeed the subject engages in any misbehavior. These notifications indicate that the individual has been denied permission and also that misbehavior has occurred.

(iv) Misbehavior

Subjects engage in misbehavior inside the data exchange network. This is termed misbehavior when a client submits too many authorizations. As a consequence, the system's trustworthiness is severely harmed. Misbehavior also has a significant impact on the system's performance and reliability. When an individual entity sends many queries, the entire network traffic was absorbed by that client alone.

(v) Procedure of SC

The procedure of the SC is illustrated by the following stages and Figure 4 depicts the design of SC.

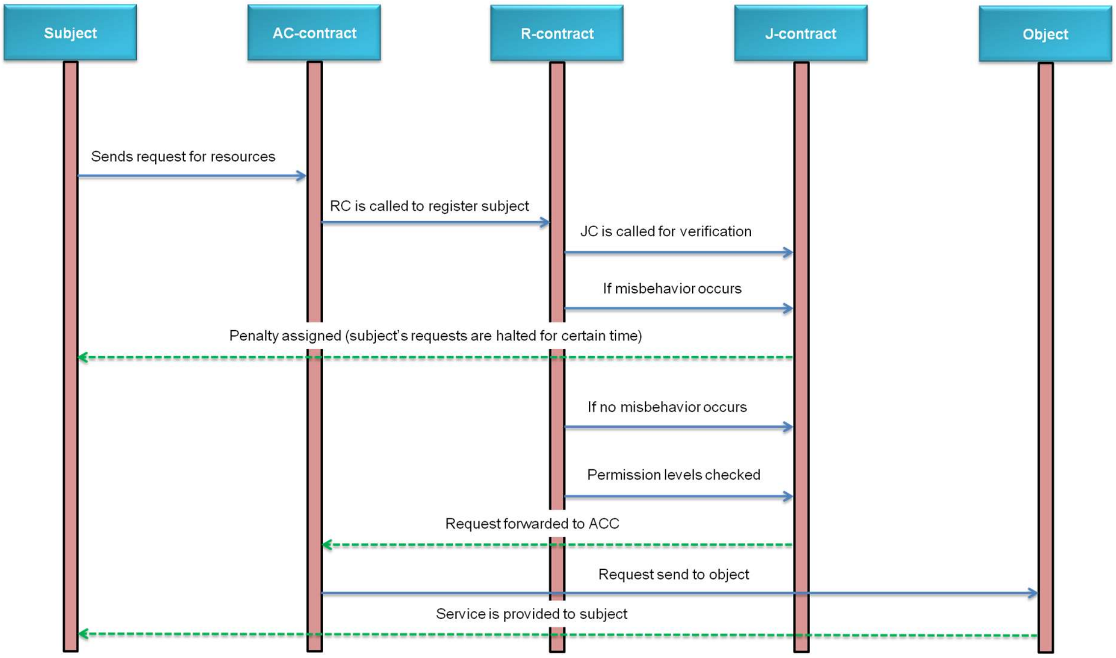
Subject
AC-contract
R-contract
J-contract
Object
Sends request for resources
RC is called to register subject
JC is called for verification
If misbehavior occurs
Penalty assigned (subject's requests are halted for certain time)
If no misbehavior occurs
Permission levels checked
Request forwarded to ACC
Request send to object
Service is provided to subject

Figure 4: Framework of Smart Contract

Step-1: A subject first transmits a message for whichever function (i.e., information, file, memory unit).

Step-2: Furthermore, SCs throughout the blockchain govern subject interaction as well as authorization.

Step-3: Whenever a subject's query is received, the AC-Contract (primary SC) gets run to govern the authorization process.

Step-4: Afterwards, the R-contract authenticates the clients by registering their data to ensure client data through a registry table.

Step-5: The platform's trustworthiness would then be managed via J-contract that examines a subject's activity.

Step-6: If indeed the subject commits in either form of misbehaviour, J-contract assesses a punishment and suspends the subject's status inside the system.

Step-7: In some other example, when no misbehaviour occurs, the subject's authorization level was examined.

Step-8: The AC-contract subsequently transmits the subjects' requests to the appropriate objects.

Step-9: Afterwards, the object fulfils the subject's requirement.

Step-10: Finally, the transaction that takes place inside the blockchain.

E. Precedence Partition Scheduling Algorithm (PPSA)

This involves selecting the sequence wherein tasks must be completed, and ensuring that scheme objectives are achievable well within specified time frame. Scheduling strategy should also observe operational constraints. Computing technologies also influence scheduling policies. The task collection would be scheduled using the contribution factor as well as the precedence relationship. A task's contribution factor (CF) indicates it could accomplish all or some of its operations. To determine cumulative performance, every task's probability is calculated as well as its completion level maintained. Rather than one element like rate, timeline, laxity, two components like contribution of task under specified structure & preceding tasks were addressed. Throughout this stage, tasks are allocated to unit operations. This stage is dynamic since task allocation to suitable core was decided when task set has been running based on contribution factor and preceding tasks. The PPSA approach is developed since it is not a globalized schedule type and tasks could be interrupted on one processor unit and resumed upon other. The proposed technique does not apply in the case of workload partitioning solutions.

During execution, based on previous tasks, a choice is made as to which core the task would be completed. In practical cases, the scheduling algorithm was forced to meet a deadline. Priority related restrictions are imposed as tasks communicate with one another to produce the logical output. The precedence relationships are depicted in Figure 5.

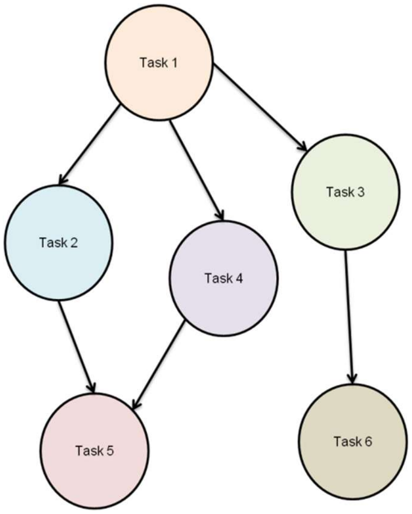

Figure 5: Relationships of precedence between Tasks

Concerning task set precedence restrictions means that succeeding tasks may not be completed until & unless corresponding previous tasks are completed. Relationships of precedence between Tasks are depicted by Directed Acyclic Graph (DAG) as (Precedence=V, B).

Here, V=Task set is represented by a set of vertices inside a graph. B=Set of edges. The messaging between tasks, and the priority relationships between tasks, are represented by a set of edges inside a graph.

Edges pointing towards the vertex indicate task precedence, but edges pointing away from vertex indicate that the vertex seems to be the predecessor for many other tasks. The proposed technique arranges tasks based on their deadlines and precedence restrictions.

Admission Controller classifies jobs statically, that means it does before they're performed. The categorization is based upon two parameters: the CF as well as the precedence relationship. This divides the tasks in two groups based on the CF and activities that came before them.

Assume S= task set with a specified amount of task and M= set of classes for categorizing tasks.

$S = \{s_1, s_2, s_3 \dots s_n\}$
(15)
$M = \{m_1, m_2\}$
(16)

Estimate CF for all tasks inside the task set as well as analyze the precedence relationship. This is accomplished through a task interaction cooperation model study, which indicates how many messages were forwarded to a completed job. Take N become the total amount of transmission of signals between the tasks that came before it. When one of these signals is not delivered, the amount of success will be reduced. When calculating CF, such a state was regarded.

Categorization of tasks was performed that are depicted below:

∀ task $s_i \in S$
$m_1 = \{s_i | s_i.CF = 0 \cap s_i.preceded\ tasks = INVALID\}$
(17)
$m_2 = \{s_i | s_i.CF = 0 \cap s_i.preceded\ tasks = INVALID\}$
(18)

Representations

m_i = Class's i^{th} Task
$m_n = Class's\ final\ task$
$core_1 = 1^{st}$ Core between every one of cores.
Prec=amount of preceded tasks
$s_k = Preceded\ Task.$
Setflag_m_1=Boolean Value for 1^{th} Core whether Busy or Idle.
Setpflag=Boolean Value for all Preceded Task Executed.

Algorithm-1: PPSA

```
Step-1: For ∀ task $s_i \in class\ m_1$
Step-2: $i = 0$ and $1 = 1$
Step-3: For $core_1$ to $core_n$
Step-4: Setflag_$m_1 = 0$
Stept-5: Accomplish $(s_i \rightarrow core_1)$
Step-6: setflag_$m_1 = 1$
Step-7: End for
Step-8: if $m_i \neq m_n$ then
Step-9: Rerun steps 3 to 7
Step-10: End for
Step-11: For ∀ task $s_i \in class\ m_2$
Step-12: $i = k = 0\ and\ 1 = 1$
Step-13: Setpflag=0
Step-14: For k to prec
Step-15: If total preceded tasks $s_k$ are accomplish l next
Step-16: For increase request processing performance, partitioning was accomplished by sending information across various tables or domains.
Step-17: Setpglag=1
```

```
Step-18: End for
Step-19: If setpflag=1 next
Step-20: Analyze preceded task $t_k$ with greatest execution time and accomplished at $core_1$
Step-21: If setflag_$m_1 = 0$ and setpflag=1 then
Step-22: Accomplish $(s_i \rightarrow core_1)$
Step-23: Setflag_$m_1 = 1$
Step-24: End if
Step-25: Else
Step-26: $core_{l+1}$
Step-27: Rerun steps 21to 28
Step-28: End else
Step-29: If $s_i \neq s_n$ then
Step-30: Rerun steps 13 to30
Step-31: End for
```

F. Enhanced particle swarm optimization (EPSO)

Across the course of its performance, the method computes a particular refreshing method for two specific parameters, namely the l-best position as well as the g-best position. This is performed by making two compares in a sequence: To determine the g-best position for every generation, the fitness of every particle at its present position was matched towards the fitness of many other particles inside the total population. Afterwards, this matches each particle’s various visiting locations with their present locations to determine the l-best place for every particle. Upon that foundation of equation (19), those two positions determine the updated velocity of every particle throughout the particle group. Two stochastic factors influence the impact of the two positions to modify the new particle velocity percentage, as shown in the equation.

$$c_{id} = wc_{id} + B_1 D_1 (k_{id} - y_{id}) + B_2 D_2 (k_{gd} - y_{id})$$

(19)

$$y_{id} = y_{id} + c_{id}$$

(20)

Here, c_{id}=present position, y_{id}=particle’s velocity, k_{gd}=particular example of k_{id}, B_i=acceleration factor, D_i= random number, w=is the inertia weight that prevents the velocity of particles from growing indefinitely, as well as a significant variable that influences the search outcome & convergence rate.

The goal of such factors would be to produce a randomized as well as legal impact of either place; as a result, certain exploration & a small exploitation must be introduced from time - to - time probabilistically. By accordance with equation (20), the algorithm adjusts the particle's present position toward a latest position depending upon the latest velocity. If every particle modifies its position, then new state of the PSO particle population would be built up by all of particles. The system then evaluates the degree of fitness for these particles based on their new placements. Finally, the algorithm repeats the method to obtain the global and local optimum positions for updating particle positions and assessing new particle positions. The algorithm-2 illustrates the suggested EPSO method for task scheduling issues.

Algorithm-2: EPSO algorithm

```
For each iteration
    for each dimension
        Return individuals that extend outside the search space's limits
    end for
    for each individual, estimate the present fitness score
        if present fitness score is higher than local best fitness score, then local best fitness score is equal to the present fitness value
            Position of individual with local best fitness is equal to the position of present individual
        end
        if present fitness score is higher than global best fitness score, then global best fitness score is equal to the present fitness value
            Position of individual with global best fitness is equal to the position of present individual
        end
    end for
    for each individual
        Estimate the modification in position for every dimension and refresh the position of every particle consequently
    end for
end
```

G. Message Digest-5 (MD-5) hash algorithm

Possessing an input element of any length and an output element of the same length is known as hashing throughout blockchain. In other hand, using a hash seems to be the only way of resolving a blockchain computation's encryption requirements. Cryptographic hash could be used to verify the validity and integrity of many forms of data. To prevent maintaining unencrypted keys in database, it is commonly used among authentication mechanisms to check pictures, messages, as well as other forms of data.

Using the blockchain like an analogy, exchanges of various lengths have been processed through the certain hashing techniques, and everything yields a constant length result. Here, we employ the MD-5 hash algorithmic procedure. In blockchain, to change the input entity to an output entity with defined length, MD-5 hash approach is employed. It's almost impossible to eliminate hash coincidences. Any randomized sub-set of a huge range of potential keys is hashed with this algorithm. Since there are a lot of coincidences, hash tables get wasteful. No null hypotheses are allowed in a hash table. To generate an MD-5 hashing, a 128-bit pattern is encoded from a string of arbitrary length. When employing the MD5 technique to encode a single string, the outcome has always been the same: a 128-bit hash. Figure 6 and algorithm-2 depict the MD-5 hashing procedure.

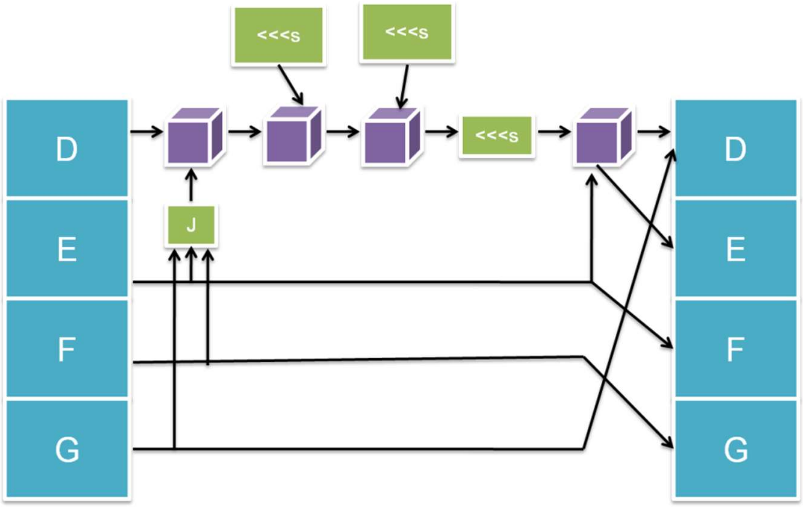

Figure 6: MD-5 hashing process

Here, the input: data with random length and output: Hashing code of 128 bits.

The 512-bit blocks are used to segment the input data (“sixteen 32-bit terms”). In cases where the data's length is indeed not divided among 512-bit blocks, the information gets padding.

There are 64 of such processes inside the MD-5 algorithm, arranged into 4 rounds of 16 processes each.

Here, J=non-linear function and every round only makes use of a single process, N_i=input data’s 32-bit block, L_i= 32-bit constant that is unique to every process, and s=shift of the left bit through s.

There are 4 32-bit terms inside the MD-5 method, which are referred classified as letters D to G. This gives the algorithm a 128-bit memory. All of them are set to the same initial values.

The state is then updated using every 512-bit message block throughout the main algorithm.

There are 4 steps in the execution of a data block known as rounds; every round refers to a non-linear function J, modular addition, and left shift. The diagram above depicts a single round of operations. In every round, one of the following J functions has been used.

$J(E,F,G)=(E \wedge F) + (\neg E \wedge G)$

$K(E,F,G)= (E \wedge G) + (F \wedge \neg G)$

$L(E,F,G)= E \text{ XOR } F \text{ XOR } G$

$M(E,F,G)= F \text{ XOR } (E + \neg G)$

Here, ‘¬’ indicates “AND”, ‘∧’ indicates “NOT”, and ‘+’ indicates “OR”.

A hash value and message digest are all terms used to describe the result.

Algorithm-3: MD-5 hashing procedure

```
Stage-1: completing the entry of data //Extends the data into K*64+56 bytes, through adding a supplement immediately following the original data. In this case, K is a number.//
Stage-2: Data length should be added
Stage-3: Initiate the variables
Stage-4: Describe the 4 32-bit variables //named D, E, F and G correspondingly.//
Stage-5: Processing of data
Stage-6: The outcome is in the order of D, E, F and G.
```

H. Delegated Proof of Stake consensus (DPoS) protocol

Between all consensus mechanisms, the DPoS would be the quickest, highest efficient, most distributed, and most versatile. Figure 6 depicts the algorithmic process. DPoS solves consensus difficulties democratically & fairly by allowing stakeholders can accept choices. Authorized agents may alter all structural features, including budgeting towards block space & transaction size. The premise is to allow every user to vote, leading in a specific number of Members who are checked as well as reported to by such super nodes on favor of the user; those super nodes have equal opportunities. DPoS is indeed a system of voting similar to that of board members. Some certain amount of super nodes is created by the coin users. As per the defined schedule, the selected nodes create blocks in turn. If a super node goes down to appropriately apply its power (for instance, by producing blocks), it would be expelled from the network, and a new super node would be chosen to substitute it. As an incentive, all representatives would obtain 10percent of the service fee included in the average level block. Payments are processed in an average with one second according to the predictable selection of block producers.

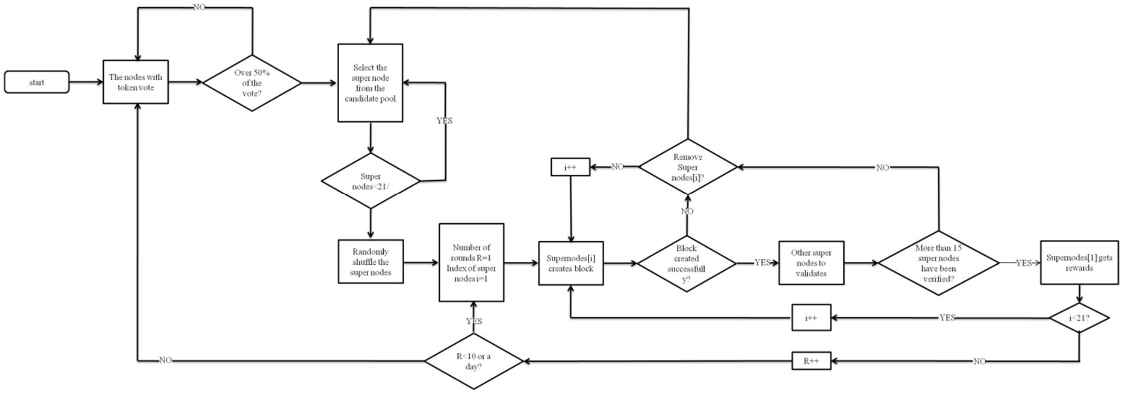
start
The nodes with token vote
Over 50% of the vote?
NO
Select the super node from the candidate pool
YES
Super nodes<21/
Randomly shuffle the super nodes
Number of rounds R=1 Index of super nodes i=1
Supernodes[i] creates block
Block created successfull y?
YES
Other super nodes to validates
More than 15 super nodes have been verified?
YES
Supernodes[1] gets rewards
i<21?
YES
i++
NO
R++
R<10 or a day?
YES
NO
Remove Super nodes[i]?
NO
i++
NO
NO

Figure 7: DPoS algorithm

When comparing to the proof of work (PoW) & proof of stake (PoS) procedures, DPoS considerably decreases the number of nodes involved for confirmation as well as accounting through preserving every participants from needless logical evaluations. The DPoS method greatly enhances performance and allows for second-level consensus confirmation. It features modest centralization rather than complete decentralization. Nevertheless, the DPoS algorithm's architecture doesn't really imply that there should be enough actual block makers. Because a single person or entity can manage several nodes, the entire system can be effectively dominated by a single entity. The f2pool, for instance, commanded 50% of the nodes in lightning bitcoin. At the same time, super node administration and commercial interests are overly centralized. If they cooperate, they will build a massive monopoly, which would be incompatible with the blockchain concept. Furthermore, the system has numerous challenges in dealing with the nodes. The establishment of some damaging nodes over time cannot be successfully prevented by community voting, posing safety hazards to the system. At the same time, super nodes decided in the case of minor set of network nodes are also not representative.

IV. RESULT AND DISCUSSION

In this phase, we are going to discuss and find out the outcomes of our proposed approach. Then these outcomes are depicted by employing the MATLAB tool. Here, we use big data approach for this investigation of performances of our proposed approach in blockchain based SCM.

Big data can be described as data resources with a massive density (large-scale), high velocity (movable), and high diversity (e.g., quantitative, textual, visual, and so on) that require cost-effective, creative kinds of data analysis enabling improved insight & strategic planning. Because of the enhanced capacity to both acquire enormous volumes of data & apply highly effective statistical tools to large sets of data, big data analytics have lately come to significant attention. Businesses' current capacity to acquire large amounts of (various) information as well as use effective methodological approaches to that data allows them to make highly complicated conclusions that were previously based (mainly or exclusively) upon human judgment as well as intuition.

The situation concerning handling vast amounts of information has remained unchanged. Information is constantly kept in big files, such as online records or complicated XML files. Every one of those files was converted to current structured information in order to meet the requirements of traditional databases. If we want to use it again, we will need to change it.

The aforementioned translation is often excessively time-consuming process, and the information kept is limited to subsequent use. To quickly extract relevant information from huge data dispersed processing takes place. Given the limited availability of several internal sources of data, it has still not been deployed. Moreover, many security protocols make integrating external data into current company IT systems more difficult. So, we concern both the scalability and security throughout this research. The performances of investigation for our proposed approach are illustrated as below mentioned.

In a private network, the company establishes its individual security procedures based on its needs. Because of its local data exchange regulation, when it is dispersed, policy has a significant impact. Various security policies might be used by a single supply chain organization to individuals in various areas. Figure 7 depicts the performance metric of security of our proposed approach and that was related to the existing approaches. The proportion of security varies with respect to distinct kinds of datasets such as structured, semi-structured, and unstructured data. Throughout this investigation, the proposed approach has the greatest rate of security level than that of the existing approaches, particularly for structured dataset over other kinds of datasets.

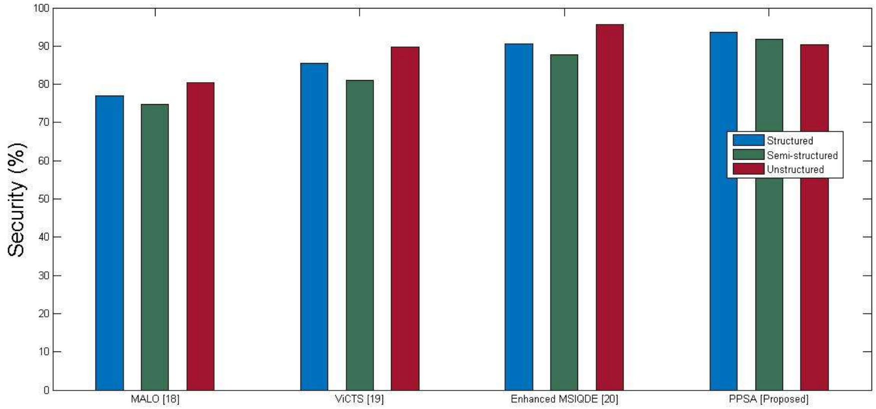

Figure 7: Comparison of security with existing and proposed approaches

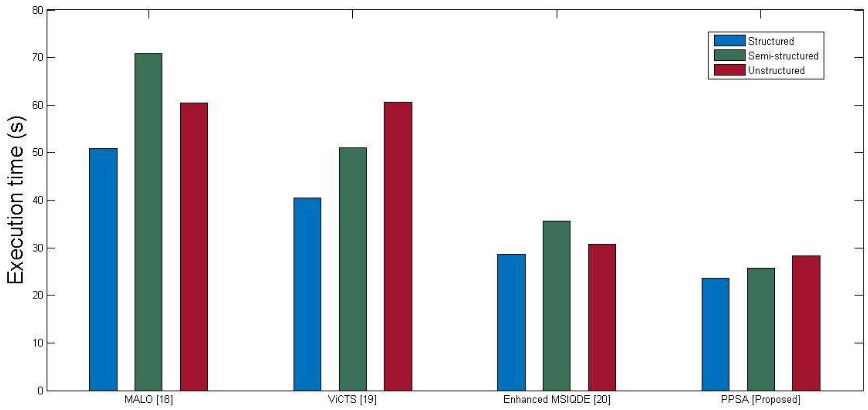

Figure 8: Comparison of execution time with existing and proposed approaches

Execution time is the time a system or functional unit takes to react to a given input. Figure 8 depicts the performance metric of execution time of our proposed approach and that was related to the existing approaches. The value of execution time varies with respect to distinct kinds of datasets such as structured, semi-structured, and unstructured data. Throughout this investigation, our proposed approach carried out the processes with lower timeslot (execution time) than that of the existing approaches. That means our proposed approach gets performs such processes with minimal time particularly for structured datasets over other kinds of datasets.

The centralized model receives requests from people all across the world, putting additional strain on servers. As a result, there's a potential that delay will increase. There would be slowness between the request as well as the reply as a result of this. Local datacenters, on either side, in dispersed network architecture, provide users with faster responses. Figure 9 depicts the performance metric of efficiency of our proposed approach and that was related to the existing approaches. The proportion of efficiency varies with respect to distinct kinds of datasets such as structured, semi-structured, and unstructured data. Throughout this investigation, the proposed approach has the greatest rate of efficiency than that of the existing approaches, particularly for structured dataset over other kinds of datasets.

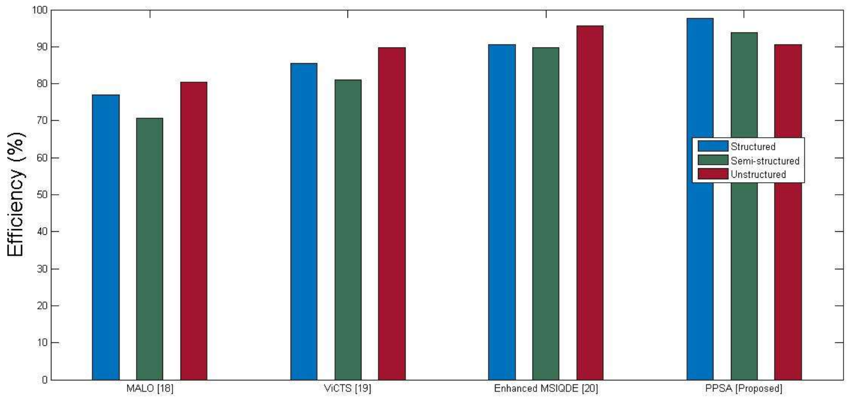

Figure 9: Comparison of efficiency with existing and proposed approaches

How much information may be transferred in a certain timeframe is called system throughput. Figure 10 depicts the performance metric of throughput (Mbps) of our proposed approach and that was related to the existing approaches. The degree of throughput varies with respect to distinct kinds of datasets such as structured, semi-structured, and unstructured data. Throughout this investigation, the proposed approach has the greatest throughput rate (Mbps) than that of the existing approaches, particularly for structured dataset over other kinds of datasets.

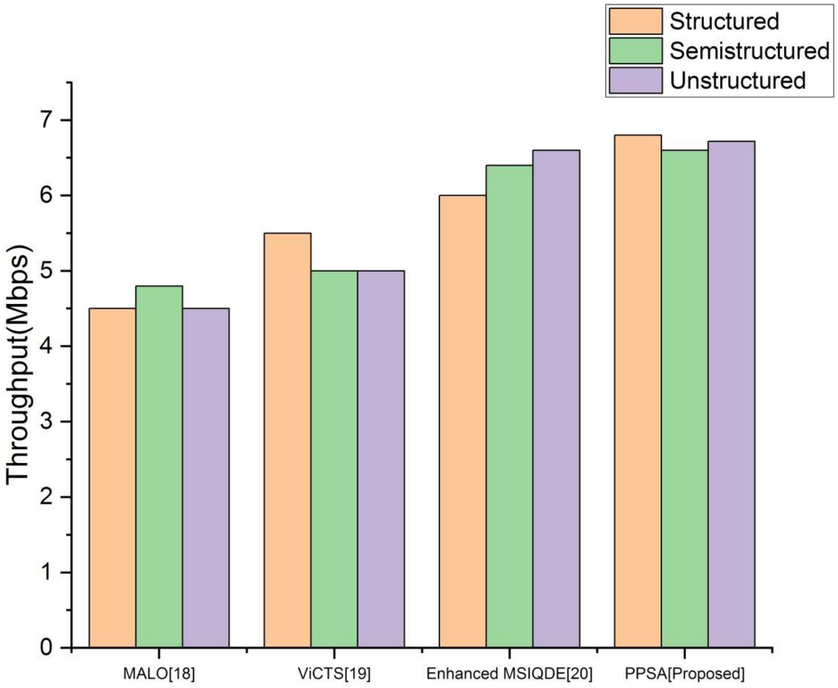

Figure 10: Comparison of throughput (Mbps) with existing and proposed approaches

Response time is the time consumed between the user's request and the application's response. In other words, the whole time taken to react to a requirement is known as the response time. Figure 11 depicts the performance metric of response time (s) of our proposed approach and that was related to the existing approaches. The response time varies with respect to distinct kinds of datasets such as structured, semi-structured, and unstructured data. Throughout this investigation, the proposed approach has the lowest response time (s) than that of the existing approaches, particularly for structured dataset over other kinds of datasets.

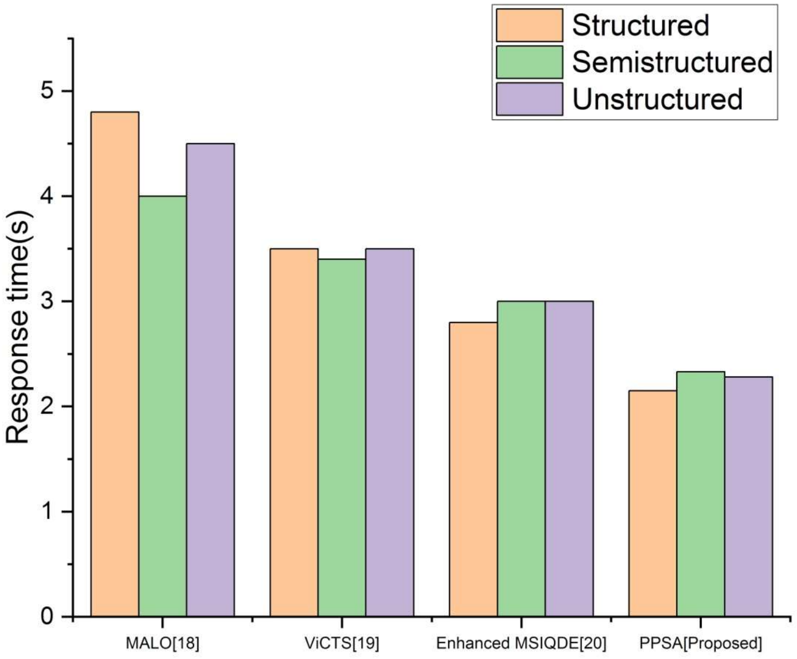

Figure 11: Comparison of response time (s) with existing and proposed approaches

As a computer's processing resources are used, or as the volume of effort a CPU handles, CPU utilization is measured. There are many factors that influence how much CPU is actually being used. Due to non-CPU resource needs, some processes demand a lot of CPU time but others do not. Figure 12 depicts the performance metric of CPU usage of our proposed approach and that was related to the existing approaches. The utilization of CPU varies with respect to distinct kinds of datasets such as structured, semi-structured, and unstructured data. Throughout this investigation, the proposed approach has the greatest CPU utilization than that of the existing approaches, particularly for structured dataset over other kinds of datasets.

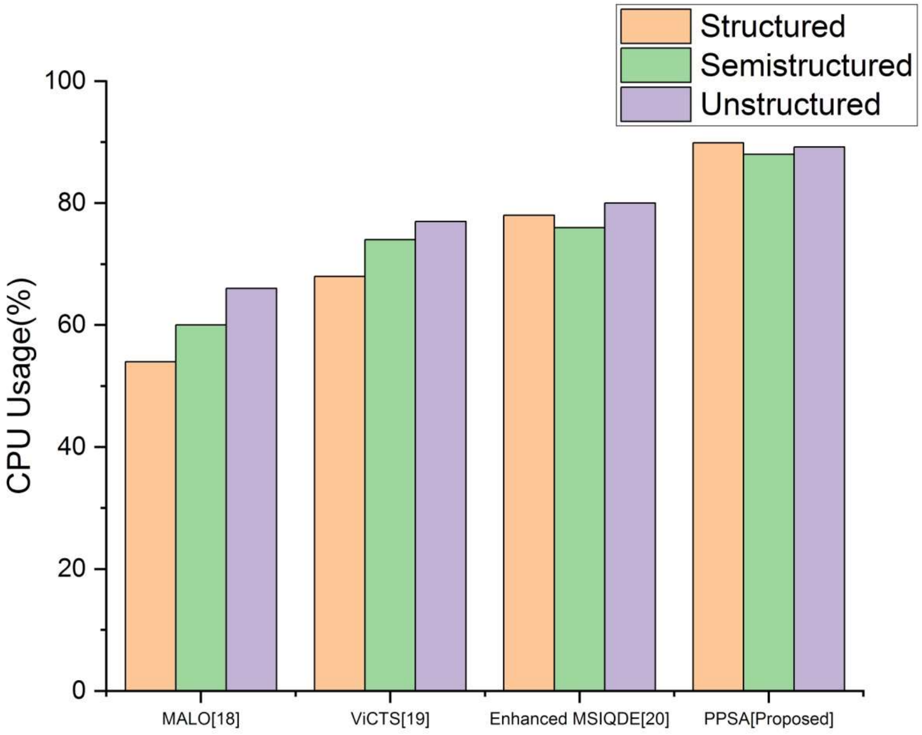

Figure 12: Comparison of CPU usage (%) with existing and proposed approaches

The total amount of energy spent by an implementation seems to be the total of the electricity utilized by each individual resource and the energy expended by the interactions in between functions. Figure 13 depicts the performance metric of energy usage of our proposed approach and that was related to the existing approaches. The utilization of energy varies with respect to distinct kinds of datasets such as structured, semi-structured, and unstructured data. Throughout this investigation, the proposed approach has the lowest energy utilization than that of the existing approaches, particularly for structured dataset over other kinds of datasets.

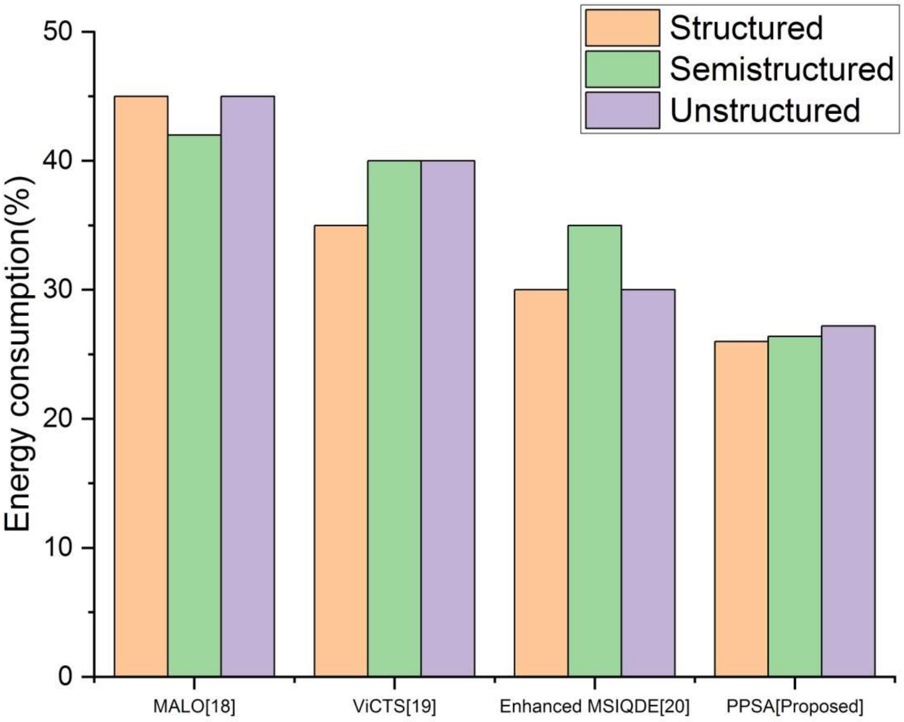

Figure 13: Comparison of energy consumption (%) with existing and proposed approaches

V. CONCLUSION

Throughout such existing researches, the scalability and protection are essential concerns for SCM. In this research, we present a unique PPSA approach for scalability optimization with secured storage of database for blockchain-based SCM of big data. For this research's investigation, we collect the distinct kinds of datasets like structured, semi-structured, and unstructured dataset. Initially, we utilize the normalization approach for pre-processing phase. Then we take the PCA approach for extracting the characteristics and deploy the Smart Contracts for authorization phase. To changing the input data into output data with defined size inside the blockchain, the MD-5 hashing procedure is employed. After that, we employ the proposed approach for enhance the storage's security level. For optimize the scalable performance, we apply EPSO approach. Then obtained dataset are stored inside the blockchain network and we utilize the DPoS protocol in the process of checking of nodes with minimal execution duration. Our proposed research accomplishes the greatest throughput, CPU usage, security, and efficiency, and minimized execution time, response time, energy consumption over the existing approaches.

REFERENCE

1. Wamba, S.F. and Queiroz, M.M., 2020. Blockchain in the operations and supply chain management: Benefits, challenges and future research opportunities. International Journal of Information Management, 52, p.102064.
2. Longo, F., Nicoletti, L., Padovano, A., d'Atri, G. and Forte, M., 2019. Blockchain-enabled supply chain: An experimental study. Computers & Industrial Engineering, 136, pp.57-69.
3. Rejeb, A., Keogh, J.G. and Treiblmaier, H., 2019. Leveraging the internet of things and blockchain technology in supply chain management. Future Internet, 11(7), p.161.
4. Sundarakani, B., Ajaykumar, A. and Gunasekaran, A., 2021. Big data driven supply chain design and applications for blockchain: An action research using case study approach. Omega, 102, p.102452.
5. Udokwu, C., Kormiltsyn, A., Thangalimodzi, K. and Norta, A., 2018, November. The state of the art for blockchain-enabled smart-contract applications in the organization. In 2018 Ivannikov Ispras Open Conference (ISPRAS) (pp. 137-144). IEEE.
6. Caro, M.P., Ali, M.S., Vecchio, M. and Giaffreda, R., 2018, May. Blockchain-based traceability in Agri-Food supply chain management: A practical implementation. In 2018 IoT Vertical and Topical Summit on Agriculture-Tuscany (IOT Tuscany) (pp. 1-4). IEEE.
7. Lou, M., Dong, X., Cao, Z. and Shen, J., 2021. SESCF: a secure and efficient supply chain framework via blockchain-based smart contracts. Security and Communication Networks, 2021.
8. Iftekhar, A., Cui, X., Hassan, M. and Afzal, W., 2020. Application of blockchain and Internet of Things to ensure tamper-proof data availability for food safety. Journal of Food Quality, 2020.
9. Thakur, S. and Breslin, J.G., 2020. Scalable and secure product serialization for multi-party perishable good supply chains using blockchain. Internet of Things, 11, p.100253.
10. Perboli, G., Musso, S. and Rosano, M., 2018. Blockchain in logistics and supply chain: A lean approach for designing real-world use cases. Ieee Access, 6, pp.62018-62028.
11. Yun, J., Goh, Y. and Chung, J.M., 2020. DQN-Based Optimization Framework for Secure Sharded Blockchain Systems. IEEE Internet of Things Journal, 8(2), pp.708-722.
12. Gao, Z., Xu, L., Chen, L., Zhao, X., Lu, Y. and Shi, W., 2018. CoC: A unified distributed ledger based supply chain management system. Journal of Computer Science and Technology, (2), pp.237-248.
13. Zhang, K. and Jacobsen, H.A., 2018. Towards Dependable, Scalable, and Pervasive Distributed Ledgers with Blockchains (Technical Report).
14. Azzaoui, A.E., Kim, T.W., Pan, Y. and Park, J.H., 2021. A Quantum Approximate Optimization Algorithm Based on Blockchain Heuristic Approach for Scalable and Secure Smart Logistics Systems. HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES, 11.

15. Li, D., Han, D., Crespi, N., Minerva, R. and Sun, Z., 2021. Fabric-SCF: A Blockchain-based Secure Storage and Access Control Scheme for Supply Chain Finance. arXiv preprint arXiv:2111.13538.
16. Westerkamp, M., Victor, F. and Küpper, A., 2020. Tracing manufacturing processes using blockchain-based token compositions. Digital Communications and Networks, 6(2), pp.167-176.
17. Chen, M., Liu, H., Wei, S. and Gu, J., 2018. Top managers' managerial ties, supply chain integration, and firm performance in China: A social capital perspective. Industrial Marketing Management, 74, pp.205-214.
18. Abualigah, L. and Diabat, A., 2021. A novel hybrid antlion optimization algorithm for multi-objective task scheduling problems in cloud computing environments. Cluster Computing, 24(1), pp.205-223.
19. Yin, D., Wang, S. and Ouyang, Y., 2020. ViCTS: a novel network partition algorithm for scalable agent-based modeling of mass evacuation. Computers, Environment and Urban Systems, 80, p.101452.
20. Deng, W., Xu, J., Gao, X.Z. and Zhao, H., 2020. An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems. IEEE Transactions on Systems, Man, and Cybernetics: Systems.