International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST2293104

Implementation of SHA256 Algorithm for Securing Medical Data using Machine Learning and Block chain Techniques

Ravindra Changala, M. Naresh
M.Tech Student, CSE Department, Newton's Institute of Engineering, Macherla, India
HOD & Associate Professor, CSE Department, Newton's Institute of Engineering, Macherla, India

Article Info
Volume 9, Issue 3
Page Number : 486-494

Publication Issue
May-June-2022

Article History
Accepted : 01 June 2022
Published : 09 June 2022

ABSTRACT

Data is the input for various artificial intelligence (AI) algorithms to mine valuable features, yet data in Internet is scattered everywhere and controlled by different stakeholders who cannot believe in each other, and usage of the data in complex cyberspace is difficult to authorize or to validate. As a result, it is very difficult to enable data sharing in cyberspace for the real big data, as well as a real powerful AI. In this paper, we propose the SecNet, an architecture that can enable secure data storing, computing, and sharing in the large-scale Internet environment, aiming at a more secure cyberspace with real big data and thus enhanced AI with plenty of data source, by integrating three key components: Blockchain-based data sharing with ownership guarantee, which enables trusted data sharing in the large-scale environment to form real big data; AI-based secure computing platform to produce more intelligent security rules, which helps to construct a more trusted cyberspace; Trusted value-exchange mechanism for purchasing security service, providing a way for participants to gain economic rewards when giving out their data or service, which promotes the data sharing and thus achieves better performance of AI. Moreover, we discuss the typical use scenario of SecNet as well as its potentially alternative way to deploy, as well as analyze its effectiveness from the aspect of network security and economic revenue.

Keywords: SecNet, Block chain, machine learning, AI, cyber security, Secure hashing, SHA

I. INTRODUCTION

The development of information technologies, the trend of integrating cyber, physical and social (CPS) systems to a highly unified information society, rather than just a digital Internet, is increasing. In such an information society, data is the asset of its owner, and its usage should be under the full control of its owner, although this is not the common case. Given data is undoubtedly the oil of the information society, almost every big company want to collect data as much as possible, for their future competitiveness. An

increasing amount of personal data, including location information, web-searching behaviour, user calls, user preference, is being silently collected by the built-in sensors inside the products from those big companies, which brings in huge risk on privacy leakage of data owners. Moreover, the usage of those data is out of control of their owners, since currently there is not a reliable way to record how the data is used and by who, and thus has little methods to trace or punish the violators who abuse those data.

That is, lack of ability to effectively manage data makes it very difficult for an individual to control the potential risks associated with the collected data. For example, once the data has been collected by a third party (e.g., a big company), the lack of access to this data hinders an individual to understand or manage the risks related to the collected data from him. Meanwhile, the lack of immutable recording for the usage of data increases the risks to abuse them. If there is an efficient and trusted way to collect and merge the data scattered across the whole CPS to form real big data, the performance of artificial intelligence (AI) will be significantly improved since AI can handle massive amount of data including huge information at the same time, which would bring in great benefits and even makes AI gaining the ability to exceed human capabilities in more areas.

II. SYSTEM ARCHITECTURE

SecNet is built as an architecture for a more secure cyberspace, by integrating three key components: Blockchain-based data sharing with ownership guarantee; AI-based secure computing platform based on big data to produce intelligent and dynamic security rules; Trust value exchange mechanism for purchasing security services. Figure 2.1.1 illustrates the overall architecture of SecNet. Nodes in SecNet are connected with Blockchain-based Networking.

In the network, nodes communicate with each other and reach a consensus based on blockchain techniques. In the meanwhile, they cooperate through the execution of smart contracts. In order to reach a consensus, either on node state or smart-contract execution results, each node contains a blockchain ledger to sync state with other nodes.

In terms of data, SecNet nodes are equipped with the data storage module and access control module for data security. SecNet nodes also have an Operation Support System (OSS) module which enables AI-based secure computing (ASC) for generating knowledge and secure rules from data.

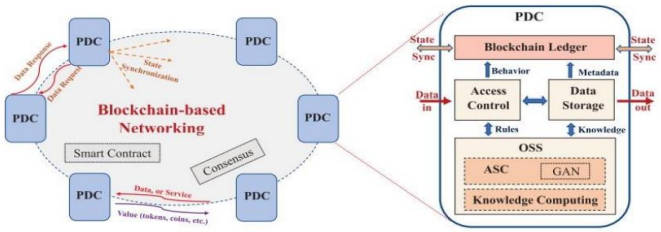

Fig 1. The Secnet Architecture

III. IMPLEMENTION OF SECNET

The traditional way of medical data management is inefficient for building a global health care system. The lack of trust mechanisms for data provenance, auditing and control, makes the sharing of valuable data impossible. Moreover, patients have to collect their medical records by themselves and then provide them to different institutions, although these medical records may be stored several times in other institutes before, because different institutions cannot easily share medical records due to no standard format for data or no economic incentive. On the other hand, medical data carries its owner's privacy information, but unfortunately, patients are in fact lack of authority for the usage of these data. SecNet employs blockchain-based data sharing guaranteeing, Smart contracts to regulate the interactions between trust-

less entities, AIbased secure computing for behaviour analysing, to effectively provide data provenance, auditing and control, as well as behaviour tracking, via a tamper-proof way.

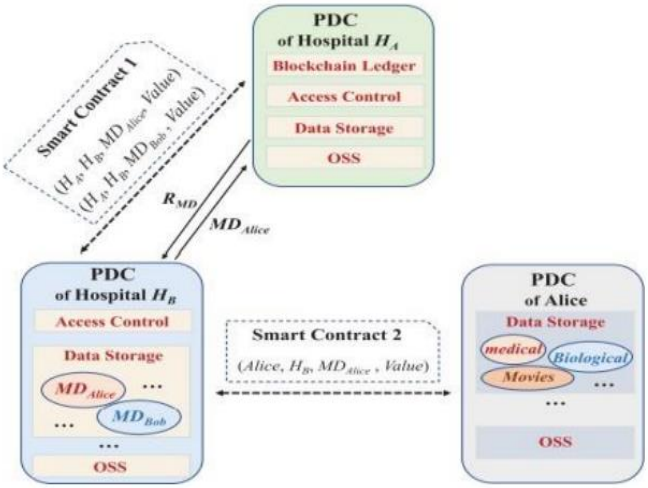

Fig 2. Medical Data Sharing Using SecNet

IV. LITERATURE SURVEY

Different authors have done their research work in the field of Data security on different domains and proved good efficiency and accuracy in removing security threats by different techniques. Some of the authors mainly concentrated on Data Security for achieving better accuracy.

Hyper connected network - A Decentralized trusted computing and networking paradigm: The development of the Internet of Things, a complex CPS system has emerged and is becoming a promising information infrastructure. In the CPS system, the loss of control over user data has become a very serious challenge, making it difficult to protect privacy, boost innovation, and guarantee data sovereignty. In this article, we propose HyperNet, a novel decentralized trusted computing and networking paradigm, to meet the challenge of loss of control over data. HyperNet is composed of the intelligent PDC, which is considered as the digital clone of a human individual; the decentralized trusted connection between any entities based on blockchain as well as smart contract; and the UDI platform, enabling secure digital object management and an identifier-driven routing mechanism. HyperNet has the capability of protecting data sovereignty, and has the potential to transform the current communication-based information system to the future data-oriented information society.

Lightweight RFID protocol for medical privacy protection in IoT: Traditional medical privacy data are at a serious risk of disclosure, and many related cases have occurred over the years. For example, personal medical privacy data can be easily leaked to insurance companies, which not only compromises the privacy of individuals, but also hinders the healthy development of the medical industry. With the continuous improvement of cloud computing and big data technologies, the Internet of Things technology has been rapidly developed. Radio frequency identification (RFID) is one of the core technologies of the Internet of Things. The application of the RFID system to the medical system can effectively solve this problem of medical privacy. RFID tags in the system can collect useful information and conduct data exchange and processing with a back-end server through the reader. The whole process of information interaction is mainly in the form of cipher text. In the context of the Internet of Things, the paper presents a lightweight RFID medical privacy protection scheme. The scheme ensures security privacy of the collected data via secure authentication. The security analysis and evaluation of the scheme indicate that the protocol can effectively prevent the risk of medical privacy data being easily leaked.

Amber - Decoupling user data from Web Applications: User-generated content is becoming increasingly common on the Web, but current web applications isolate their users' data, enabling only restricted sharing and cross-service integration. We

believe users should be able to share their data seamlessly between their applications and with other users. To that end, we propose Amber, an architecture that decouples users' data from applications, while providing applications with powerful global queries to find user data. We demonstrate how multi-user applications, such as e-mail, can use these global queries to efficiently collect and monitor relevant data created by other users. Amber puts users in control of which applications they use with their data and with whom it is shared, and enables a new class of applications by removing the artificial partitioning of users' data by application.

Enhancing selectivity in big data: Today's companies collect immense amounts of personal data and enable wide access to it within the company. This exposes the data to external hackers and privacy-transgressing employees. This study shows that, for a wide and important class of workloads, only a fraction of the data is needed to approach state-of-the-art accuracy. We propose selective data systems that are designed to pinpoint the data that is valuable for a company's current and evolving workloads. These systems limit data exposure by setting aside the data that is not truly valuable.

H.S. Jennath et al., discovers the possibility of forming trusted blockchain based Artificial Intelligence models in e-Health sector which can provide a transparent platform for consent-based data sharing. Training and testing of AI or machine learning models is not performed on similar datasets. For training, authorized data sets are employed which further helps in prediction.

Tharukarupasinghe et al., has recommended a dynamic consent management architecture based upon the technique of blockchain technology and smart contracts that complies six prime design targets. Also, there is no need of the patients for the data verification process. There are three smart contracts involves namely: 1) registration, 2) request policy and 3) response policy.

RenpengZou et al., has suggested the framework of medical data sharing and privacy preserving eHealth system that employs the blockchain technique and integrates Repucoin with the SNARKs based chameleon hash function for preventing against potential blockchain threats.

V. PROPOSED SYSTEM

In the proposed system we use the concept called Private Data Centers (PDC) with lockchain and AI technique to provide security to user's data. In this technique 3 functions will work which describe below

Blockchain: Blockchain-based data sharing with ownership guarantee, which enables trusted data sharing in the large-scale environment to form real big data. In this technique users can define access control which means which user has permission to access data and which user cannot access data and Blockchain object will be generate on that access data and allow only those users to access data which has permissions. In Blockchain object user will add/subscribe share data and give permission.

Artificial Intelligence: AI-based secure computing platform to produce more intelligent security rules, which helps to construct more trusted cyberspace. AI work similar to human brain and responsible to execute logic to check whether requesting user has permission to access shared data. If access is available then AI allow Blockchain to display share data otherwise ignore request.

Machine Learning: Machine learning (ML) is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the extract information from the

data. In that case, we apply machine learning. With the abundance of datasets available, the demand for machine learning is in rise. Many industries apply machine learning to extract relevant data. The purpose of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves without being explicitly programmed. Many mathematicians and programmers apply several approaches to find the solution of this problem which are having huge data sets.

Table 1. ML algorithms for various model building approaches

Learning type	Model building	Examples
Supervised	Algorithms or models learn from labelled data (task-driven approach)	Classification, regression
Unsupervised	Algorithms or models learn from unlabeled data (Data-Driven Approach)	Clustering, associations, dimensionality reduction
Semi-supervised	Models are built using combined data (labelled + unlabeled)	Classification, clustering
Reinforcement	Models are based on reward or penalty (environment-driven approach)	Classification, control

Rewards: In this technique all users who is sharing the data will earn rewards point upon any user access his data. trusted value-exchange mechanism for purchasing security service, providing way for participants to gain economic rewards when giving out their data or service, which promotes the data sharing and thus achieves better performance of AI.

VI. MEDICAL DATA

Patient: Patients first create his profile with all disease details and then select desired hospital with whom he wishes to share/subscribe data. While creating profile application will create Blockchain object with allowable permission and it will allow only those hospitals to access data.

Hospital: Hospital1 and Hospital2 are using in this application as two organizations with whom patient can share data. At a time, any hospital can login to application and then enter search string as disease name.

The eHealth system is a major contribution in development of the overall cloud environment by outsourcing their critical data management systems. Such eHealth framework helps in building bonds among the major entities viz, patients, doctors and hospital, which also leads to generation of voluminous amount of eHealth information. Primarily, the eHealth records comprise of diagnostic records, health monitoring data, medical histories and prescriptions as indicated by [1]. It's highly crucial that all the preserved eHealth information is kept highly confidential and protected, failing which can result in misleading and wrong treatment thus causing a negative effect on patient's life.

Certain prime issue to be confronted are of authorized access, interoperability of health records, rule enforcement or data sharing policies pertaining to the shared content and securely sharing eHealth records

as per [2]. Today, achieving patient's data confidentiality stands as the most critical challenge that remains of paramount importance for maintaining data integrity. With peak in digitization of healthcare system, there is significant improvement in precise analysis, patient's overall care as well as safe and secure eHealth data accessing. Caseless increase in eHealth data along with its processing and storing can definitely impact the scalability aspect of the eHealth system. Attending this concerning issue is out of reach for the old conventional approaches. Towards this, the cloud technology assures to be the most effective means in achieving the storage and processing of such massive amount of data indicates [3].

VII. IMPLEMENTATION

Blockchain seems complicated, and it definitely can be, but its core concept is really quite simple. A blockchain is a type of database. For the purpose of understanding blockchain, it is instructive to view it in the context of how it has been implemented by Bitcoin. Like a database, Bitcoin needs a collection of computers to store its blockchain. For Bitcoin, this blockchain is just a specific type of database that stores every Bitcoin transaction ever made. In Bitcoin's case, and unlike most databases, these computers are not all under one roof, and each computer or group of computers is operated by a unique individual or group of individuals.

In this model, blockchain is used in a decentralized way. However, private, centralized blockchains, where the computers that make up its network are owned and operated by a single entity, do exist. In a blockchain, each node has a full record of the data that has been stored on the blockchain since its inception. For Bitcoin, the data is the entire history of all Bitcoin transactions. If one node has an error in its data it can use the thousands of other nodes as a reference point to correct itself. This way, no one node within the network can alter information held within it. Because of this, the history of transactions in each block that make up Bitcoin's blockchain is irreversible.

If one user tampers with Bitcoin's record of transactions, all other nodes would crossreference each other and easily pinpoint the node with the incorrect information. Thissystem helps to establish an exact and transparent order of events. For Bitcoin, this information is a list of transactions, but it also is possible for a blockchain to hold a variety of information like legal contracts, state identifications, or a company's product inventory.

VIII. SHA 256 ALGORITHM

SHA 256 is a part of the SHA 2 family of algorithms, where SHA stands for Secure Hash Algorithm. Published in 2001, it was a joint effort between the NSA and NIST to introduce a successor to the SHA 1 family, which was slowly losing strength against brute force attacks.

The significance of the 256 in the name stands for the final hash digest value, i.e. irrespective of the size of plaintext/cleartext, the hash value will always be 256 bits. The other algorithms in the SHA family are more or less similar to SHA 256. Now, look into knowing a little more about their guidelines.

SHA256 Algorithm

Input: Block of Message

Output: Fixed Size bits

Step 1: Pre-Processing

i). Indexing and Padding with 0's until data is a multiple of 512, less 64 bits.

Step 2: Initialize Hash Values

ii). Now create hash values

Step 3: Initialize Round Constants

iii). Similar to step 2, we are creating some constants. This time, there are 64 of them.

Step 4: Chunk Loop
IV). The following steps will happen for each 512-bit "chunk" of data from our input.
Step 5: Create Message Schedule
V). Copy the input data from step 1 into a new array where each entry is a 32-bit word
Step 6: Compression
vi). Initialize variables and set them equal to the current hash values respectively
Step 7: Modify Final Values
vii). after the compression loop, but still, within the chunk loop, we modify the hash values by adding their
respective variables to them.
Step 8: Concatenate Final Hash

The SHA-256 algorithm is one flavor of SHA-2, which was created by the National Security Agency in 2001 as a successor to SHA-1. SHA-256 is a patented cryptographic hash function that outputs a value that is 256 bits long. In encryption, data is transformed into a secure format that is unreadable unless the recipient has a key. In its encrypted form, the data may be of unlimited size, often just as long as when unencrypted.

In hashing, by contrast, data of arbitrary size is mapped to data of fixed size. For example, a 512-bit string of data would be transformed into a 256-bit string through SHA-256 hashing. In cryptographic hashing, the hashed data is modified in a way that makes it completely unreadable. It would be virtually impossible to convert the 256- bit hash mentioned above back to its original 512-bit form. So why would you want to create a scrambled message that can't be recovered? The most common reason is to verify the content of data that must be kept secret. For example, hashing is used to verify the integrity of secure messages and files. The hash code of a secure file can be posted publicly so users who download the file can confirm they have an authentic version without the contents of the file being revealed. Hashes are similarly used to verify digital signatures.

The US government requires its agencies to protect certain sensitive information using SHA-256. It is built with a Merkle-Damgård structure derived from a one-way compression function itself created with the Davies-Meyer structure from a specialized block cipher.

Padding Bits

It adds some extra bits to the message, such that the length is exactly 64 bits short of a multiple of 512. During the addition, the first bit should be one, and the rest of it should be filled with zeroes.

Fig 3.Totsl length to be 64 bits less than the multiples of 512

Padding Length

You can add 64 bits of data now to make the final plaintext a multiple of 512. You can calculate these 64 bits of characters by applying the modulus to your original clear text without the padding.

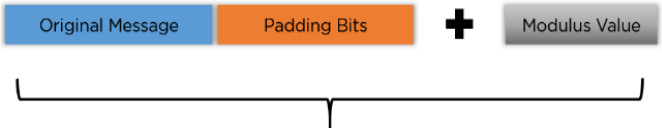

Fig 4. Final data to be hashed as a multiples of 512

Compression Functions

The entire message gets broken down into multiple blocks of 512 bits each. It puts each block through 64 rounds of operation, with the output of each block

serving as the input for the following block. The entire process is as follows:

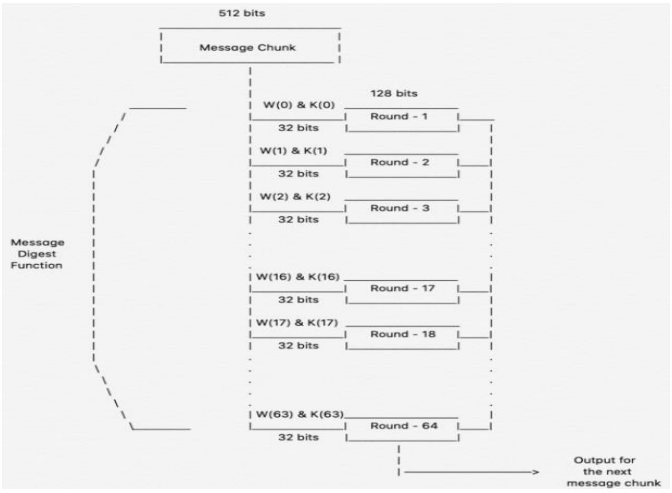

Fig 5. Compression function

Initialising the Buffers:

You need to initialize the default values for eight buffers to be used in the rounds as follows:

```
a = 0x6a09e667    k[0..63] :=
b = 0xbb67ae85       0x428a2f98, 0x71374491, 0xb5c0fbcf, 0xe9b5dba5, 0x3956c25b, 0x59f111f1, 0x923f82a4, 0xab1c5ed5,
c = 0x3c6ef372       0xd807aa98, 0x12835b01, 0x243185be, 0x550c7dc3, 0x72be5d74, 0x80deb1fe, 0x9bdc06a7, 0xc19bf174,
d = 0xa54ff53a       0xe49b69c1, 0xefbe4786, 0x0fc19dc6, 0x240calcc, 0x2de92c6f, 0x4a7484aa, 0x5cb0a9dc, 0x76f988da,
e = 0x510e527f       0x983e5152, 0xa831c66d, 0xb00327c8, 0xbf597fc7, 0xc6e00bf3, 0xd5a79147, 0x06ca6351, 0x14292967,
f = 0x9b05688c       0x27b70a85, 0x2e1b2138, 0x4d2c6dfc, 0x53380d13, 0x650a7354, 0x766a0abb, 0x81c2c92e, 0x92722c85,
g = 0x1f83d9ab       0xa2bfe8a1, 0xa81a664b, 0xc24b8b70, 0xc76c51a3, 0xd192e819, 0xd6990624, 0xf40e3585, 0x106aa070,
h = 0x5be0cd19       0x19a4c116, 0x1e376c08, 0x2748774c, 0x34b0bcb5, 0x391c0cb3, 0x4ed8aa4a, 0x5b9cca4f, 0x682e6ff3,
                     0x748f82ee, 0x78a5636f, 0x84c87814, 0x8cc70208, 0x90befffa, 0xa4506ceb, 0xbef9a3f7, 0xc67178f2
```

You also need to store 64 different keys in an array, ranging from K[0] to K[63]. They are initialized as follows:

Output

With each iteration, the final output of the block serves as the input for the next block. The entire cycle keeps repeating until you reach the last 512-bit block, and you then consider its output the final hash digest. This digest will be of the length 256-bit, as per the name of this algorithm.

With the SHA 256 algorithm being implemented thoroughly since the early 90s, there are specific applications that you can look into. You will see them in the next section.

IX. THE BENEFITS OF SHA-256

We use SHA-256 because this 256-bit key is much more secure than other common hashing algorithms. a hash value is a unique number string that's created through an algorithm, and that is associated with a particular file. If the file is altered in any way, and you recalculate the value, the resulting hash will be different. In other words, it's impossible to change the file without changing the associated hash value as well. Without going into too much technical detail, here are the key benefits of SHA-256:

- It's a secure and trusted industry standard: SHA-256 is an industry standard that is trusted by leading public-sector agencies and used widely by technology leaders.
- Collisions are incredibly unlikely: There are 2256 possible hash values when using SHA-256, which makes it nearly impossible for two different documents to coincidentally have the exact same hash value. (More on this in the following section).
- The Avalanche effect: Unlike some older hashing algorithms, even a very minor change to the original information completely changes the hash value what is known as an avalanche effect.

X. CONCLUSION

In order to leverage AI and blockchain to fit the problem of abusing data, as well as empower AI with the help of blockchain for trusted data management in trust-less environment, we propose the SecNet, which is a new networking paradigm focusing on secure data storing, sharing and computing instead of communicating. SecNet provides data ownership guaranteeing with the help of blockchain technologies, and AI-based secure computing platform as well as blockchain-based incentive mechanism, offering paradigm and incentives for data merging and more powerful AI

to finally achieve better network security. Moreover, we discuss the typical use scenario of SecNet in medical care system, and gives alternative ways for employing the storage function of SecNet. Furthermore, we evaluate its improvement on network vulnerability when countering DDoS attacks, and analyse the inventive aspect on encouraging users to share security rules for a more secure network.

XI. FUTURE ENHANCEMENT

In future work, we will explore how to leverage blockchain for the access authorization on data requests, and design secure and detailed smart contracts for data sharing and AI-based computing service in SecNet. In addition, we will model SecNet and analyse its performance through extensive experiments based on advanced platforms i.e., integrating IPFS [27] and Ethereum [28] to form a SecNet-like architecture which would result in better security with huge advancement.

XII. REFERENCES

[1]. Hyperconnected network: A decentralized trusted computing and networking paradigm - IEEE Netw., vol. 32, pp. 112–117, January 2018.

[2]. Lightweight RFID protocol for medical privacy protection in IoT - IEEE Trans Ind. Informat., vol. 14, no. 4, pp. 1656–1665, April 2018.

[3]. Enhancing selectivity in big data - IEEE Security Privacy, vol. 16, no. 1, pp.34–42, January2018.

[4]. OpenPDS: Protecting the privacy of metadata through SafeAnswers - PLoSONE, vol. 9, no. 7, 2014, Art. no. e98790.

[5]. End-to-end privacy for open big data markets - IEEE Cloud Comput., vol. 2no. 4, pp. 44–53, April 2015.

[6]. Adaptable blockchain-based systems: A case study for product traceability -IEEE Soft, vol 34, no. 6, pp. 21–27, November 2017.

[7]. Ethereum: A secure decentralized generalized transaction ledger – Ethereum Project Yellow Paper, June 5, 2019

Cite this article as :

Ravindra Changala, M. Naresh, "Implementation of SHA256 Algorithm for Securing Medical Data using Machine Learning and Block chain Techniques", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 3, pp. 486-494, May-June 2022. Available at doi : https://doi.org/10.32628/IJSRST2293104
Journal URL : https://ijsrst.com/IJSRST2293104