IEEE Open Journal of
Vehicular Technology

Received 5 June 2023; accepted 11 September 2023. Date of publication 22 September 2023;
date of current version 16 October 2023. The review of this article was coordinated by Editor Peter Han Joo Chong.

Digital Object Identifier 10.1109/OJVT.2023.3317640

A Lightweight Stochastic Blockchain for IoT Data Integrity in Wireless Channels

SUSHILA DHAKA [1] (Graduate Student Member, IEEE), YU-JIA CHEN [2] (Senior Member, IEEE), SWADES DE [3] (Senior Member, IEEE), AND LI-CHUN WANG [1] (Fellow, IEEE)

[1] Department of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
[2] Department of Communication Engineering, National Central University, Taoyuan 320, Taiwan
[3] Department of Electrical Engineering, Indian Institute of Technology Delhi, Delhi 110016, India

CORRESPONDING AUTHOR: LI-CHUN WANG (e-mail: lichun@g2.nctu.edu.tw)

This work was supported in part by the National Science and Technology Council under Grants MOST 110-2221-E-A49-039-MY3, MOST 111-2221-E-A49-071-MY3, NSTC 111-2634-F-A49-010, and NSTC 111-3114-E-A49-001, Taiwan, in part by the Center for Open Intelligent Connectivity from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan, and in part by the Higher Education Sprout Project of the National Yang Ming Chiao Tung University and Ministry of Education (MOE), Taiwan.

ABSTRACT Trustworthy validators selection is crucial as validators determine whether a block should be added to its chain. In this article, we proposed a novel confidence score-based lightweight stochastic blockchain for wireless Internet of Things (IoT) systems. The received signal strength is used to define the trust level of a wireless IoT node to facilitate the selection of more trustworthy validator nodes, thereby reducing the possibility of selecting malicious validators. We used a lightweight authentication protocol called Burrows-Abadi-Needham (BAN) logic to prevent unauthorised information leakage. A formal security analysis of BAN logic is provided for proving secure and fresh data storage. Our analysis and simulations reveal that the probability of successful defense against data integrity attacks (i.e., the probability that the majority of the validator nodes are not compromised) can be improved up to two times higher than that associated with the closest competitive scheme of random selection in stochastic blockchains. Our results further reveal that the probability of successful defense depends on the total number of nodes in the network and the number of validator nodes. The proposed blockchain concept can be easily implemented in various wireless IoT environments to enhance the successful defense of the system for maintaining IoT data integrity.

INDEX TERMS Blockchain, data integrity, Internet of Things (IoT), network security.

I. INTRODUCTION

Wireless Internet of Things (IoT) technology connects numerous devices and sensors through a distributed wireless network environment. The collected IoT data can facilitate intelligent decision-making. Wireless sensor networks (WSNs) is the integral part of the IoT to facilitate the wireless interconnections of things [1]. Here, our research is focused on the security of the resource-constrained wireless-connected IoT devices. However, this technology's wireless characteristics complicate the secure exchange of IoT data among heterogeneous IoT devices [2]. The untrusted communication environment may cause leakage of data [3], an authentication mechanism is required for identifying the communicating party and avoiding spoofing hijackers. Therefore lightweight authentication scheme needs to be addressed for low-powered IoT devices. A major security issue in IoT networks is data integrity. Conventional data verification methods that rely on a trusted central entity are not suitable for distributed IoT systems. In addition, IoT devices are usually resource-limited [4], precluding the implementation of complex data verification algorithms. The aforementioned security challenges can be resolved by implementing suitable lightweight authentication techniques and integrating wireless IoT and blockchain. In a blockchain network, transaction records are stored as blocks, each of which contains the hash value of the previous block to which it was linked [5]. Any change in a block results in a change in the corresponding hash, resulting in an immutable chain. Notably, blockchains can achieve data consistency among nodes through a consensus mechanism. The preservation of data integrity in wireless IoT relies on leveraging the chain structure and consensus mechanism. In this regard, trustworthy validator nodes verify the data consistency via the consensus mechanism prior to its inclusion in the chain. Moreover, the immutable characteristics inherent in blockchain technology serve as a protective barrier, impeding any unauthorized modifications to the stored data.

II. RELATED WORK

The authors of [6] proposed a lightweight blockchain-based secure distributed key management scheme for flying ad hoc networks (FANET). The authors of [7] surveyed blockchain for securing vehicular networks. The coexistence of heterogeneous networks in an unlicensed spectrum by introducing blockchain implementation with proof of strategy consensus mechanism is presented by [8]. The majority of the papers on blockchain-enabled wireless networks analyzed the latency, scalability, and throughput of the system. There are very few studies [9] regarding the successful defense of blockchain against malicious attacks. Some studies have integrated blockchains into wireless IoT by using various consensus mechanisms, such as Proof of Work (PoW) [10] and Proof of Stake (PoS) [11]. The main limitation therein is the high computational cost of block mining that became a bottleneck when there is more data transfer in IoT compared to traditional cryptocurrency scenarios. Specifically, each block appended to PoW and PoS mechanisms must be verified by all the nodes within the blockchain network. This is not practical for resource-limited IoT devices [12]. An alternative to PoW and PoS protocols is the Practical Byzantine Fault Tolerance (PBFT) algorithm, in which only a few preselected validators are required to reach a consensus mechanism [13]. Its throughput and storage efficiency are superior to those of PoW and PoS protocols, and the computational costs are relatively low. This comes at the cost of security, however; the attack tolerance is less than 33% [14].

The author of [15] proposed the concept of fast and secure consortium blockchains with lightweight block verifiers (LBVs). LBVs are edge devices that help typical miners in verifying the blocks. To provide high data integrity, the selection of trustworthy miner nodes is also important. The author of [16] proposed the validator selection technique for integrating blockchains into drones in 5G. Validators are selected based on their interaction frequency, and direct and indirect opinions from other drones. The author of [17] proposed a consensus algorithm for blockchain-based IoT that selects one master node among all nodes based on voting, and then the master node selects a few validator nodes for verifying the data instead of wasting the resource in the competition of becoming validators. As a result, this scheme may not be very secure, as there is the possibility of an attack on the master node.

The author of [18] proposed a trust-based privacy-preserving scheme for IoT networks for improving cooperative sensing. Trust is adaptive in nature based on nodes' historical and current performance. Data is stored in blockchain for maintaining immutability. Trust is an important factor to be considered while dealing with data integrity. For the calculation of the trustworthiness of a node or confidence of a node, there are a few parameters [19] that are commonly used: direct trust and indirect trust, historical behaviour and current behaviour, and adaptive trust based on the current behaviour and historical behaviour of the node. There are very less researchers who exactly talk about the exact parameters defining trust. In [20], the concept of age of information is introduced to verify the data freshness. Further, the author of [21] analyzed the impact of timely updates of information in the blockchain.

Therefore, maintaining authenticated, fresh, and trustworthy data are important factors for maintaining data integrity. While considering blockchain-IoT integration in wireless scenarios, wireless channel characteristics play an important role. To the best of the authors' knowledge, there is no research for validator selection that considers wireless characteristics and stochastic nature. An efficient trust-based lightweight consensus mechanism is needed [22] for maintaining data integrity in blockchain-enabled IoT.

A. MOTIVATION

An authentication scheme is required for nodes to prevent unauthorized access to data from external attackers. We employed BAN logic authentication owing to its lightweight nature and suitability for resource-constrained IoT devices [3], [23]. Furthermore, low block overhead and trustworthiness of validator nodes are essential to ensure high data integrity in blockchain-IoT environments. Since IoT devices are typically resource-limited, nodes cannot compete for block validation. An attacker can manipulate the blockchain by compromising the majority of the validator nodes because they perform block mining. Our previous work [9] proposed a stochastic blockchain network with multiple randomly selected nodes as validator nodes. We demonstrated the data integrity with low block verification overhead by introducing randomness in validator selection. However, this work does not consider reliability between IoT nodes, device authentication, and data freshness; and each node has an equal chance of being elected as a validator node. In this article, we discuss preventing unauthorized access to data and maintaining the freshness of data using BAN logic. Further, we propose a novel trust-aware validator selection scheme to reduce the possibility that compromised nodes are selected as validator nodes. By implementing weighted validator selection, the consensus mechanism becomes lightweight and efficient.

B. RESEARCH CONTRIBUTIONS

The main contributions of this research are summarized as follows:

- Our novel method, which can be used in the stochastic selection of trusted block validators, is based on estimates of the confidence scores of IoT nodes. The confidence score of a node can be calculated by comparing the strength of the signal it receives with its reported location. In our design, each node has a probability of

being selected as a validator on the basis of its confidence score.
- The probability of successful defense, defined as the probability that the number of compromised validators is less than half of the total number of validators, is analyzed. The impacts of the number of nodes and the number of validators on the security performance under varying levels of attacker ability are also analysed.
- Extensive simulations in various network attack scenarios have been performed. The validator selection scheme outperformed the random selection scheme.
- Based on design goals of high data integrity, Burrows-Abadi-Needham (BAN) logic-based data authentication is done for proving data security and freshness. Mathematical analysis of BAN logic in the considered scenario is also proposed.

The remainder of the article is organized as follows. In Section III, we present a background on blockchains. In Section IV, our solution is described. Section V presents the mathematical analysis of the proposed scheme. Section VI presents the performance evaluation, and the conclusions are provided in Section VII.

III. BACKGROUND ON BLOCKCHAINS

The data in a blockchain is stored in the form of blocks after verification by validator nodes and data storage is distributed in nature [24]. Each block contains the previous block's hash value, which makes the blockchain immutable as any data changes affect both current and subsequent blocks. Based on controlling authority, blockchain may be broadly classified as a public blockchain, private blockchain, and consortium blockchain. In a public blockchain, there is no central authority, which is open to everyone [25] like bitcoin [26]. However, the private blockchain is managed by a single organization with full control of validator selection. Hyperledger Fabric [27] managed by the Linux foundation is the most common open-source platform for supporting private blockchains. The consortium blockchain is a hybrid blockchain that is controlled by a group of validator nodes. It is suitable for heterogeneous IoT systems with various administrative domains [28].

Depending on the required security level and the network environment, different consensus mechanisms are used by validator nodes to verify the blocks. PoW, PoS and PBFT are the most commonly used consensus mechanism. In PoW, nodes compete to solve a computational puzzle; the node that solves the puzzle first is rewarded. PoS is intended to solve the problem of high energy consumption in PoW. The validator nodes in the PoS mechanism are selected on the basis of the value of coins held (i.e., the stake). The probability of being selected as a validator is determined by the nodes' respective stakes. In case of malicious behavior, the nodes are punished and their stakes are reduced [29]. In PBFT, the consensus of the new block is reached if and only if no less than two-thirds of validators confirm the block within a given time period [30]. This is intended to reduce transaction time and increase network scalability.

This article considers the consortium blockchain because it is more suitable for resource-constrained IoT devices. As a consensus mechanism, PoW requires high computational power that is not compatible with resource-constrained IoT devices. In PoS protocols, the probability of a node being selected as a validator is positively correlated with the value of the stake it holds. The public nature of stake information enables the prediction of which nodes participate in the block validation process. To resolve this security vulnerability, in our previous work [9], we introduced the stochastic consensus mechanism. We demonstrated that the randomness introduced during the validator selection process significantly can reduce the attack success probability. Validator nodes are responsible for verifying the block data as well as maintaining the data integrity. Therefore, selecting a trustworthy node is an important issue for blockchain, especially in the open wireless communication scenario. Because the validator selection mechanism proposed in [9] involved uniform probability, it is therefore inherently unable to reflect the node heterogeneity. Hence, in this article, we propose a stochastic weighted selection of validators for IoT data integrity using the calculated confidence score based on wireless characteristics.

IV. SYSTEM MODEL

The system model has three types of nodes: sensor nodes, cluster head nodes, and validator nodes. Let S $= \{S_1, S_2, S_3, \ldots\ldots, S_n\}$ be the set of randomly distributed N sensor nodes. Sensor nodes have low computational power and sense the target continuously. The target T, which may either be the transmitter to be localized or any primary user, is under continuous detection by the sensor nodes. Sensor nodes with high confidence scores are selected as validator nodes. The sensor nodes transmit data, together with the corresponding sensor's location (Loc) information as well as the received signal strength indicator (RSSI), to their designated cluster head (CH). Cluster head nodes are IoT edge nodes having higher computational energy than sensor nodes.

CH receives data (RSSI, Loc) from its nearby sensors and calculates the confidence score of each sensor node. Further, CH transfers data to the base station (BS), also known as the destination node. The destination node (D) is the highly secure node and selects validator nodes stochastically based on their weight. Further, validator nodes send their data to the smart contract (SC) and majority-based data is selected for blockchain (BC) storage. The system architecture of our proposed system is shown in Fig. 1.

Sensor model: After detecting a target, the sensor node sends (RSSI, Loc) to the nearest cluster head. The IoT edge nodes act as the cluster heads are assigned with sensors. A sensor associated with a particular cluster head is based on the shortest Euclidean distance among the nearby cluster heads. Based on the data obtained from sensors, IoT edge nodes make an estimation of the target's position with a certain error d_{err} (Fig. 2). As shown in Fig. 2, an annulus corresponds to

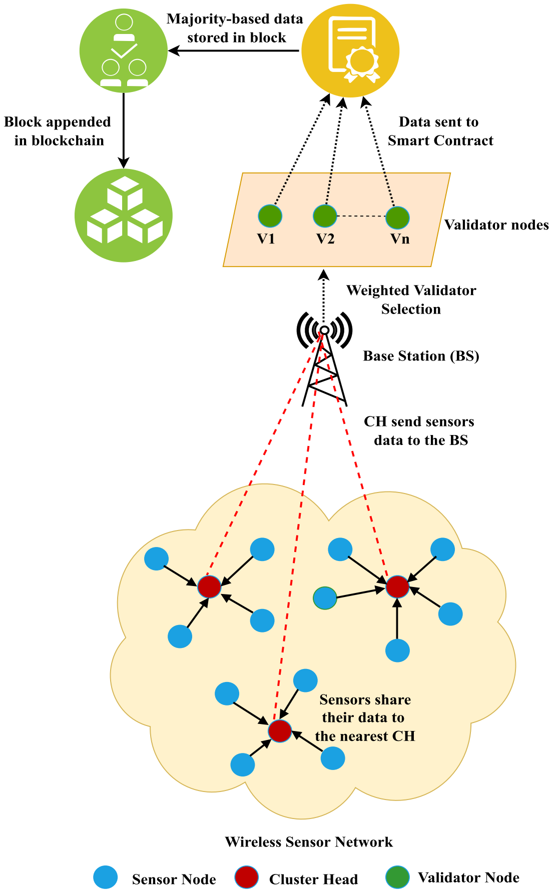

Majority-based data stored in block
Block appended in blockchain
Data sent to Smart Contract
V1
V2
Vn
Validator nodes
Weighted Validator Selection
Base Station (BS)
CH send sensors data to the BS
Sensors share their data to the nearest CH
Wireless Sensor Network
Sensor Node
Cluster Head
Validator Node

FIGURE 1. Blockchain-IoT system architecture.

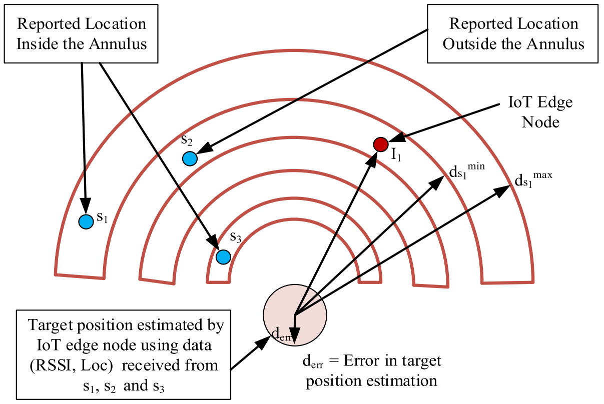

Reported Location Inside the Annulus
Reported Location Outside the Annulus
IoT Edge Node
s_2
I_1
ds_1^{min}
ds_1^{max}
s_1
s_3
d_{err}
Target position estimated by IoT edge node using data (RSSI, Loc) received from s_1, s_2 and s_3
d_{err} = Error in target position estimation

FIGURE 2. Estimation of sensor position.

each sensor node. The lower and upper limits of the annulus represent the minimal and maximal approximated distances of the sensor, respectively. The thickness of the annulus represents the certainty regarding the position of the sensor zone.

Blockchain model: The blockchain considered here is a consortium blockchain wherein the data read and write operations are controlled by designated validator nodes. The probability of a node being selected as a validator node depends on its confidence scores. Therefore, the blockchain consensus is very similar to the PoS one. Instead of making a stake-based validator selection, we are making a trust-based validator selection. It utilizes the resource inequality flaw of PoS advantageously as inequality in validator node selection probabilities. The final data selection for block mining is performed according to the majority-based selection of the validator nodes. Each node's confidence score and time stamp are stored in the blockchain.

Threat model: Any node that attempts to temper the data or inject malicious data is considered an attacker. Any number of nodes can be randomly attacked [9]. This depends on the attack capability C_A, which can be defined as the number of nodes compromised in a single attempt. If a sensor node is tempered in an attack, it presents falsified data (RSSI, Loc). As confidence score of each sensor node is calculated based on its data using log-distance path loss model. The IoT edge node knows the true position of the target and can estimate the position of the sensor node. The higher value of the thickness of the annulus reduces the confidence score and consequently reduces the probability of being selected as validators. Herein, we increase our rate of successful defense through a strategy change.

Protocol: Our system comprises four key phases of activity.

- Sensing phase: Sensors detect the target and report to the nearest IoT edge node.
- Weight assignment phase: The confidence score of a node corresponds to its probability of selection. The weight of each node is directly proportional to this probability.
- Validator selection phase: The higher is the weight assigned to the node, the greater is its likelihood of being selected as a validator.
- Blockchain phase: The block is mined according to a majority-based data selection process and is broadcasted to all the nodes for state updates.

A. PROPOSED CONFIDENCE SCORE BASED WEIGHT ASSIGNMENT

The goal of weight assignment is to identify the truthfulness of the sensor node in a distributed manner by using fundamental sensor-reported information. IoT edge node I_j estimates the position of the target by using the log-distance path loss model. Localization methods [31] include trilateration and multilateration. Trilateration determines the node position by using the intersection of three circles of three anchor nodes. Hence, more than three nodes are required for localization. If the distance measurements are noisy, the accuracy of position estimation is compromised. Multilateration requires distance measurement from more than three nodes. The author of [32] explains various localization techniques based on distance, the angle of arrival, and the time of arrival. However, to maintain system simplicity and energy efficiency, we used a log-distance path loss model. The estimated position of the target may vary because of one or both of the following points:

1) the falsification of sensor data by malicious sensors
2) model noise and other inaccuracies

Algorithm 1: Proposed Confidence Score Based Weight Assignment Algorithm.

```
1:  Function Weight assignment Input (map, [RSSI, Loc])
2:  $d_0 =$ Diameter(map)
3:  $[Loc_T, d_{err}] =$ Distributed target localization
4:  $d_{I_j} =$ Euclidean distance $(Loc_T, Loc_{I_j})$
5:  $d_{s_i} =$ Euclidean distance $(Loc_T, Loc_{s_i})$
6:  for k=1:N do
7:  Estimate annulus by calculating $[d_{s_i}^{\min}, d_{s_i}^{\max}]$
8:  Calculate confidence score as $C_{s_i} = \frac{1-(d_{s_i}^{\max}-d_{s_i}^{\min})}{d_0}$
9:  if $C_{s_i} \geq C_{s_{th}}$ then
10:   $\frac{V}{N} \times C_{s_i} + k$
11: else $P_i = \frac{1}{N}$
12: Assign weight of each node as $W_i = \frac{P_i}{\sum P_i}$
13: Output $W_i$ for all $S_i \in S$
```

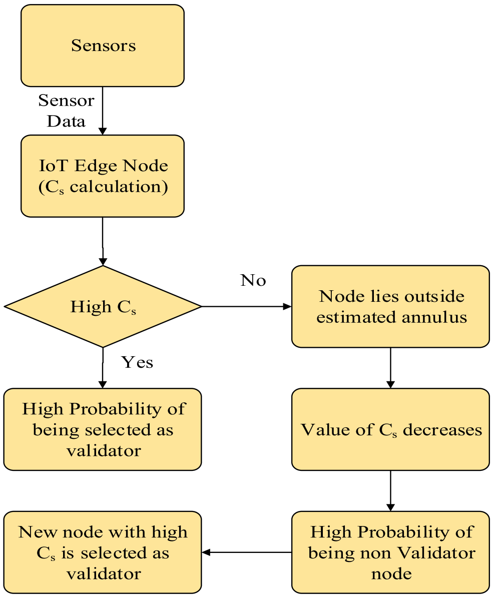

Sensors
Sensor Data
IoT Edge Node (C$_s$ calculation)
High C$_s$
No
Node lies outside estimated annulus
Yes
High Probability of being selected as validator
Value of C$_s$ decreases
New node with high C$_s$ is selected as validator
High Probability of being non Validator node

FIGURE 3. Process of validator selection.

In consideration of an allowance for error, we incorporated an error factor d_{err} into our formula; d_{err} is the error in the estimated position of the target by IoT edge nodes caused by noise or other factors such as signal distortion. The true location of the target lies within the circular region of radius d_{err} centered around the approximated target position.

The proposed weight assignment steps are presented in Algorithm 1. The distance from the target to the IoT edge node (I_j) is defined as $d_{I\,j}$ and estimated as Loc_T. The location of the IoT edge node is $Loc_{I\,j}$(line 4). Owing to uncertainty in the target location, the actual distance from the target to the IoT edge node lies in the range of $(d_{I_j} - d_{err})$, $(d_{I_j} + d_{err})$. Lines 6–13 explain the steps for weight assignment for each sensor within the set S. Regarding weight estimation, if the confidence score is greater than the threshold, the corresponding probability is calculated and the weight is assigned accordingly. Otherwise, the node is assigned 1% probability. These aforementioned nodes contain anomalies or a smaller amount of true data. Assigning a lower confidence score to malicious nodes prevents them from being the validator nodes. In the present study, we assigned a higher weight to truer nodes to maximize their likelihood of being selected as validators.

1) ESTIMATION OF THE TARGET POSITION ZONE

Using the log distance path loss model, the power of the signal transmitted from the IoT edge node and the power of the signal transmitted from the target to the sensor node is calculated. The estimated minimal and maximal distance are calculated by using the (1):

$$d_{s_i}^{\min} = \left(d_{I_j} - d_{err}\right) \times 10^{\left(\frac{P_{r,I_j} - P_{r,s_i} - x_g}{10\gamma}\right)}$$

$$d_{s_i}^{\max} = \left(d_{I_j} + d_{err}\right) \times 10^{\left(\frac{P_{r,I_j} - P_{r,s_i} - x_g}{10\gamma}\right)} \tag{1}$$

where x_g is a zero-mean Gaussian random variable that represents a shadowing effect and γ is the path loss exponent with a value that varies according to the characteristic of the area considered (e.g., rural or urban). As for the other components of the equations, d_{err} is the error in the estimated distance, d_{I_j} is the distance from the target to the IoT edge node, P_{r,I_j} is the power received from the signal transmitted from the target at the IoT edge node, and P_{r,s_i} is the power received from the signal transmitted from the target to the sensor node. Estimation of the target position zone is given by ($d_{s_i}^{\max}$- $d_{s_i}^{\min}$), which defines the annulus of sensor s_i.

2) CONFIDENCE SCORE CALCULATION AND WEIGHT ASSIGNMENT

After $d_{s_i}^{\max}$ and $d_{s_i}^{\min}$ are calculated, the position of the target can be estimated. Moreover, the confidence score can be calculated as follows:

$$C_{s_i} = \frac{1 - \left(d_{s_i}^{\max} - d_{s_i}^{\min}\right)}{d_0} \tag{2}$$

where d_0 is the reference distance used for normalization. According to the confidence score, the truthfulness of a node can be estimated; accordingly, the probability that a node is selected as a validator is determined. The greater is the confidence score C_{s_i}, the more truthful is the node, as C_{s_i} is defined on the basis of $d_{s_i}^{\max} - d_{s_i}^{\min}$, i.e. the thickness of annulus. The higher value of confidence score corresponds to the higher likelihood of being selected as a validator node. The probability of being selected as a validator P_i, a linear function similar to the linear function of confidence scores used in [33], is given by:

$$P_i = \begin{cases} \frac{V}{N} \times C_{s_i} + k, & \text{if } C_{s_i} \geq C_{th} \\ \frac{1}{N} & \text{otherwise} \end{cases} \tag{3}$$

Algorithm 2: Weight-Based Validator Selection Algorithm.

```
1: Input ($S_{id}, S_{weight}, V$)
2: Repeat step 3 and 4 for $k = 1, 2, \ldots.., v$
3: The probability of selection $S_i$ as validator is:
   $P_i(k) = \frac{W_i}{\sum_{v_j \in (S-V)} W_j}$
4: Randomly select an item $S_k \in (S - V)$ and insert to V
5: Output ($V_1, V_2, \ldots.., V_v$)
```

where N is the total number of sensor nodes in the network, V is the number of validator nodes selected, k is a constant, and C_{s_i} is the confidence score of the i_{th} node. The value of P_i may be biased by considering higher value of k. However, to avoid any bias we used $k = 0$ in our calculations. On the basis of the probability value, the weight of each node W_i is determined as follows:

$$W_i = \frac{P_i}{\sum P_i} \tag{4}$$

where $\sum P_i$ is the sum of the probability of all nodes. The weighted selection of validators is complete after all nodes have been assigned weights.

B. VALIDATOR SELECTION AND BLOCK MINING

After weight assignment, the validator nodes are selected stochastically using weighted random selection (WRS) [34] as shown in Algorithm 2 below:

In Algorithm 2, S_{id} presents the list of sensor id, S_{weight} presents the list of weights of all sensors, and V present the list of validator nodes which is initially empty. The total number of validator to be selected are presented by v, $P_i(k)$ presents the probability of k_{th} node to be selected as a validator, W_i is the weight of i_{th} sensor node. In line 1, S_{id}, S_{weight} and V are given as input to the algorithm. Steps 3 and 4 are repeated v times for selecting v validator nodes. The validator nodes are selected based on the formula given at step 3. Once a node is selected as validator node, it is removed from $S - V$ and inserted to V. For Algorithm 2, the probability that the node with weight W_n is selected as validator is $\frac{W_n}{W_1+W_2+\cdots..+W_n}$ when W_n is the first validator node to be selected. The probability of the second validator node is $\frac{W_{n-1}}{W_1+W_2+\cdots..+W_{n-1}}$, etc. as per research [34].

Further, we will discuss the validator selection strategy, system defense strategy, and block mining in blockchains, which is relatively less time and energy-consuming than other methods and provides high data security.

Validator selection: The concept behind validator selection is shown in Fig. 3. According to sensor-received data, the IoT edge node calculates the confidence score of each node. Nodes with high and low confidence scores have high and low probabilities of being selected as validators, respectively. As nodes are compromised by an attacker, the sensor node begins giving falsified data to the IoT edge node. This immediately results in confidence score reductions in all compromised nodes. As the confidence scores decrease, the estimated position of the sensor node will be outside the annulus. The weights of these nodes decrease, meaning that they are less likely to be selected as validator nodes. Hence, the IoT edge node selects new nodes with high confidence scores as validator nodes. Truer nodes have very high probabilities of selection. By selecting validator nodes with awareness of wireless channel characteristics, the data security of the system is increased.

Fig. 4 presents cases of successful and unsuccessful defense in blockchains. In the example shown in Fig. 4, three validator nodes are involved. In the successful defense case, only one validator node is compromised in the attack; the remaining validator nodes contain true data. The validator nodes give their data to the destination node and the destination node selects the data to be stored based on majority-based selection. Successful defenses in blockchains involve block mining that hinges on majority-based data selection with true data. In the unsuccessful defense case, two of the three validator nodes are compromised. In other words, the majority of the validator nodes have falsified data and, according to the majority-based selection, this falsified data is involved in block mining. After majority-based selection, the data selected by the destination node is updated in the blockchain as the blockchain is distributed and transparent in nature.

Block mining: The destination node based on the data received from the validator nodes performs it. The final selection of data for block storage is performed through majority-based selection. Owing to the selection of truer nodes as validators, the likelihood that these nodes are falsifying data is extremely low. The complete workflow of our proposed system is shown in Fig. 5.

C. RELATIONSHIP BETWEEN TRUSTWORTHINESS AND RECEIVED SIGNAL STRENGTH OF A NODE

Algorithm 1, implies that the value of the confidence score depends on the RSSI value of the sensor node. The higher value of RSSI supports the higher trustworthiness of the node. The author of [35] proposed a trust calculation method using the value of RSSI for the trust calculation of the node. Fig. 6 presents the relationship based on Neyman-Pearson Hypothesis [36]. A ROC curve illustrates the performance of a detector (binary classifier) by plotting the probability of detection (P_d) with respect to the probability of false positive (P_f) for different values of signal-to-noise ratio (SNR) [37]. The value of P_d increases and the value of P_f decreases with the increase in SNR value respectively, resulting in identifying the node behaviour. It implies that nodes with a higher SNR value results in higher trustworthiness nodes.

D. APPLICATION SCENARIOS FOR OUR PROPOSED MECHANISM

Our proposed system model shown in Fig. 1 can be applied for various IoT applications such as secure data collection, trustworthy data fusion and aggregation for cooperative sensor fusion, efficient target handover, environmental monitoring

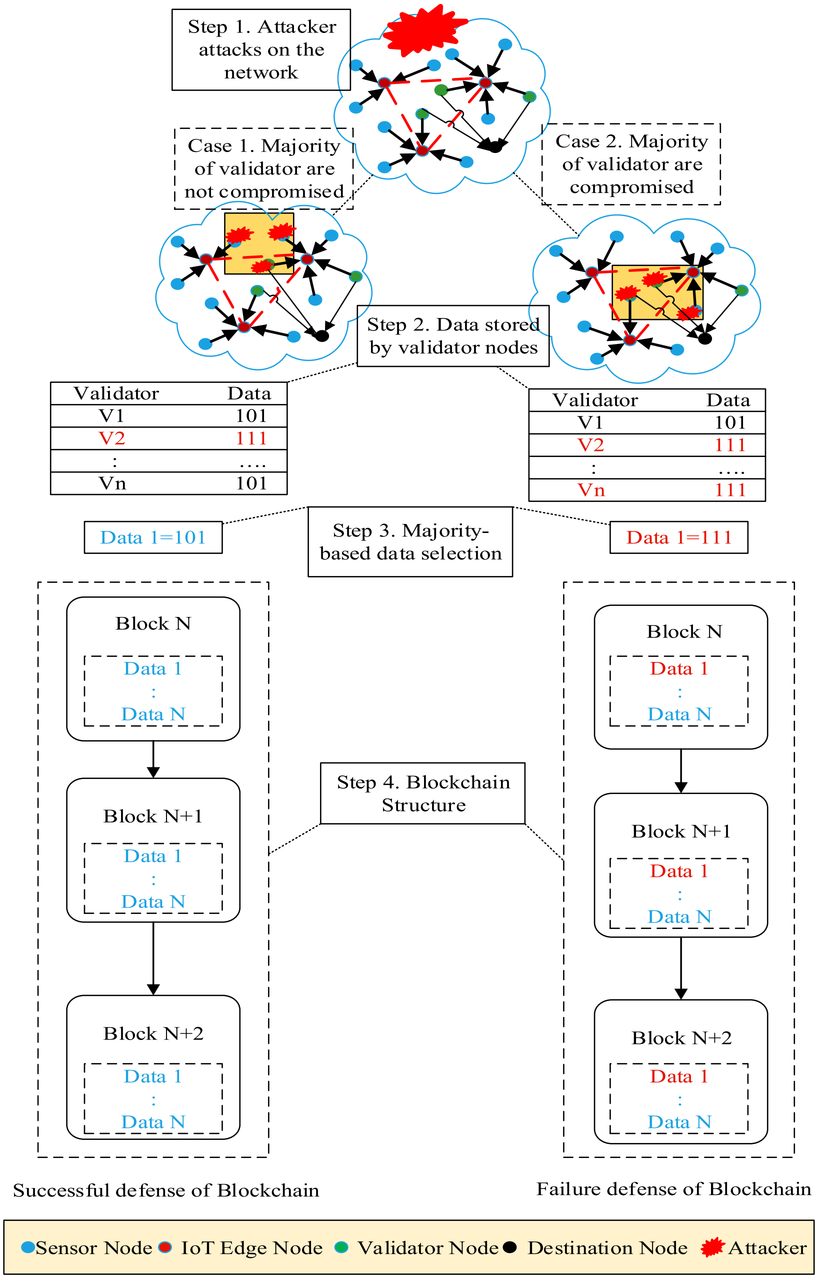

Step 1. Attacker attacks on the network
Case 1. Majority of validator are not compromised
Case 2. Majority of validator are compromised
Step 2. Data stored by validator nodes
Validator | Data
V1 | 101
V2 | 111
: | ....
Vn | 101
Validator | Data
V1 | 101
V2 | 111
: | ....
Vn | 111
Data 1=101
Step 3. Majority-based data selection
Data 1=111
Block N
Data 1
:
Data N
Block N+1
Block N+2
Step 4. Blockchain Structure
Successful defense of Blockchain
Failure defense of Blockchain
●Sensor Node ● IoT Edge Node ● Validator Node ● Destination Node Attacker

FIGURE 4. Illustration of the attacker scenario of successful defense and failure defense of blockchain with an example of V = 3.

for large agricultural area even in the presence of malicious node. The use case of secure data collection using cooperative sensor data fusion is explained below:

- In cooperative sensor data fusion, all the nodes will give their data to the corresponding CH. The cluster head assigns the weight of each sensor node based on our proposed scheme in Algorithm 1 and sends the data to the base station. The complete workflow of the system is explained in fig. 5. The final data storage is based on majority based scheme, so presence of some malicious node will not affect the data integrity. Typically, in such scenario CH are the points of interest for the attackers. However, our scheme does not use the CH for the validator process and uses weight-based validators selection. This makes our system more robust even in the presence of malicious nodes.

V. MATHEMATICAL ANALYSIS

A. FORMAL ANALYSIS USING BAN LOGIC

This section presents the formal analysis of our scheme using BAN logic [38]. The analysis aims to prove the correctness and freshness of the data stored in the blockchain. Firstly,

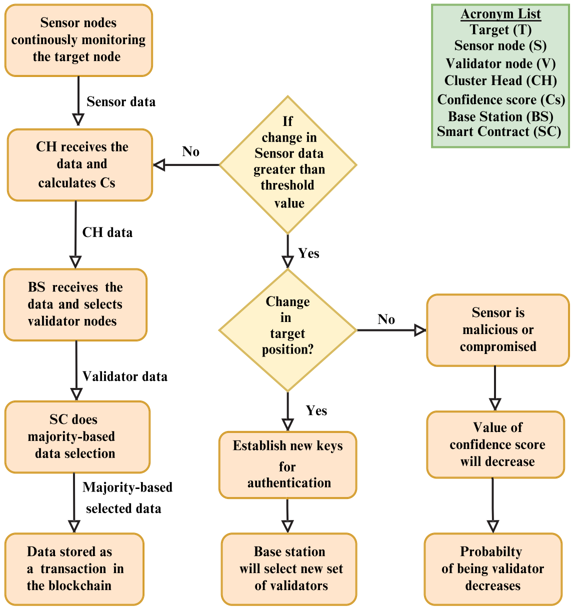

Sensor nodes continously monitoring the target node
Sensor data
CH receives the data and calculates Cs
CH data
BS receives the data and selects validator nodes
Validator data
SC does majority-based data selection
Majority-based selected data
Data stored as a transaction in the blockchain
If change in Sensor data greater than threshold value
No
Yes
Change in target position?
Yes
No
Establish new keys for authentication
Base station will select new set of validators
Sensor is malicious or compromised
Value of confidence score will decrease
Probabilty of being validator decreases
Acronym List
Target (T)
Sensor node (S)
Validator node (V)
Cluster Head (CH)
Confidence score (Cs)
Base Station (BS)
Smart Contract (SC)

FIGURE 5. Illustration of the complete workflow of the system model for authentication, key selection and validator selection.

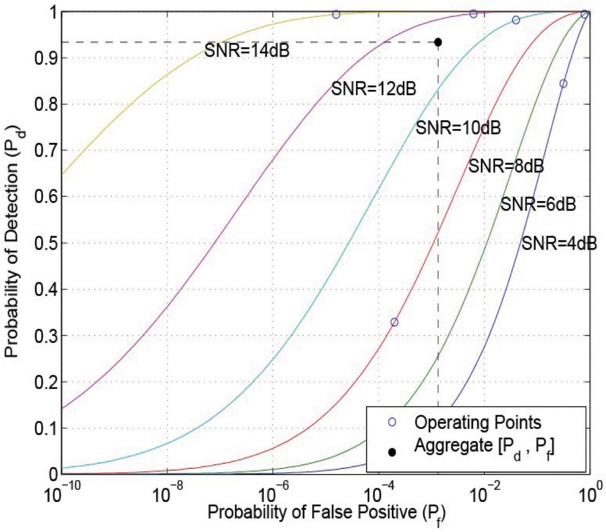


FIGURE 6. Receiver operating characteristics (RoC) [36].

we illustrate the notation and logical postulates of BAN logic.

1) BASIC NOTATIONS OF BAN LOGIC

BAN logic has its syntax and semantics for security proof. The logic considers several objects: principals, encryption keys, and formulas. The principals may be people, computers and services. The encryption keys are shared keys, public-private key pairs, session keys and secret keys based on the considered scenario. Formulas are also known as statements. We assume M and N as principals, K as the shared key, SK as the session key between principals and X as the statement. The logical description is as follows:

- $M \mid\equiv X$: Principal M believes the statement X and act as X is true.
- $M \xleftrightarrow{SK} N$: SK is the shared session key between M and N for communication.
- $M \triangleleft X$: Principal M sees statement X and can read it.
- $M \mid\sim X$: Principal M once said statement X.
- $M \Rightarrow X$: Principal M has jurisdiction over X which means M believes X and M has authority over X.
- $\#(X)$: Statement X is fresh, which implies X shared for the first time in the current run of the protocol.
- $\overset{K}{\mapsto} M$: Key K is a public key over M.
- $M \overset{K}{\leftrightarrow} N$: Principal M and N use key k for communication.
- $\{X\}_K$: It states that message X encrypted by key K.

2) LOGICAL POSTULATES

This section discusses logical postulates used in proofs using BAN logic.

- *Message meaning rule:* It concerns the interpretation of communicated messages i.e. how the principal derives belief about the origin of the message. For shared key K, the message meaning rule is postulated as follows:

$$\frac{M \mid\equiv N \overset{K}{\leftrightarrow} M, M \triangleleft [X]_K}{M \mid \equiv N \mid \sim X}.$$

  It states that if principal M believes that the key is shared with N and sees message X encrypted under the key K, then M believes that N once said X.
- *The nonce-verification:* This rule demonstrates the message's freshness and the sender still believes in the message X. It is postulated as follows:

$$\frac{M \mid\equiv \#(X), M \mid\equiv N \mid\sim X}{M \mid\equiv N \mid\equiv X}.$$

- *The jurisdiction rule:* It states that if M believes that N has jurisdiction over X, then M trusts N about the truth of statement X. The jurisdiction rule is postulated as follows:

$$\frac{M \mid\equiv N \Rightarrow X, M \mid\equiv N \mid\equiv X}{M \mid\equiv X}.$$

- *Fresh conjuncatenation rule:* If principal M believes about the freshness of X, then U also believes (X, Y) are fresh. This postulate can be represented as follows: It states that if one part of the formula is fresh, then the whole formula must be fresh. It is postulated as follows:

$$\frac{M \mid\equiv \#(X)}{M \mid\equiv \#(X, Y)}.$$

  If principal M trusts the freshness of formula X, then it also trusts the freshness of the formula (X, Y).

3) METHOD

In the BAN logic scenario, we have principals that like to communicate with each other. However, they do not trust each other. There is a server having jurisdiction over keys, and both principals believe it. The server helps principals establish trusted communication based on three major considerations:

- Verification of message origin
- Verification of message freshness
- Verification of the origin's trustworthiness

4) GOALS OF AUTHENTICATION

This section discusses the goals we want to prove using BAN logic in our scenario. BAN logic focuses on the proof of good and fresh data. In our case, we want to store true and fresh data in the blockchain. The base station (BS) performs a weighted selection of validator nodes (V) after receiving data from all the cluster heads (CH). The smart contract (SC) selects the majority-based data (X) from validator nodes for storage in the blockchain (BC). Based on the considered scenario, in this article four goals are defined as follows:

$$Goal1 : S \mid\equiv CH \mid\equiv CH \overset{K}{\leftrightarrow} S.$$

$$Goal2 : BC \mid\equiv V \mid\equiv V \overset{K}{\leftrightarrow} BC.$$

$$Goal3 : BC \mid\equiv X.$$

$$Goal4 : BC \mid\equiv \#X.$$

Goal 1 defines the trust between the cluster head and the sensor node. Goal 2 defines the trust between the blockchain and the validator node. Goal 3 defines the trust of BC on stored X and Goal 4 defines the freshness of finally stored data X in BC.

5) ASSUMPTIONS

These define the initial keys shared between protocols, principals generating a new nonce, and the trustworthiness of principals in certain ways [38]. Assumptions are always made to guarantee the success of the protocol. These assumptions act as a premise for the logic analysis. We defined eight assumptions A_1 to A_8. A_1 to A_4 are assumptions corresponding to the first set of BAN logic and A_5 to A_8 are assumptions corresponding to the second set of BAN logic. In the assumption, A_1 defines the shared key SK between CH and BS. A_2 defines the shared key (SK) between S and BS. A_3 and A_4 define the freshness of timestamps shared between $(CH,\ BS)$ and $(S,\ CH)$ respectively. A_5 and A_6 define the shared keys K_{VBS}, K_{SCBS} between $(V,\ BS)$ and $(BC,\ SC)$ respectively and have belief in keys.

$$A_1.\ CH| \equiv CH \xleftrightarrow{SK} BS. \qquad A_2.\ S| \equiv S \xleftrightarrow{SK} BS.$$

$$A_3.\ CH| \equiv \#(T_{BS}). \qquad A_4.\ S| \equiv \#(T_{CH}).$$

$$A_5.\ V| \equiv V \xleftrightarrow{K_{VBS}} BS. \qquad A_6.\ BC| \equiv BC \xleftrightarrow{K_{SCBC}} SC.$$

$$A_7.\ V| \equiv \#\,(T_{SC})\,. \qquad A_8.\ BC| \equiv \#(T_V).$$

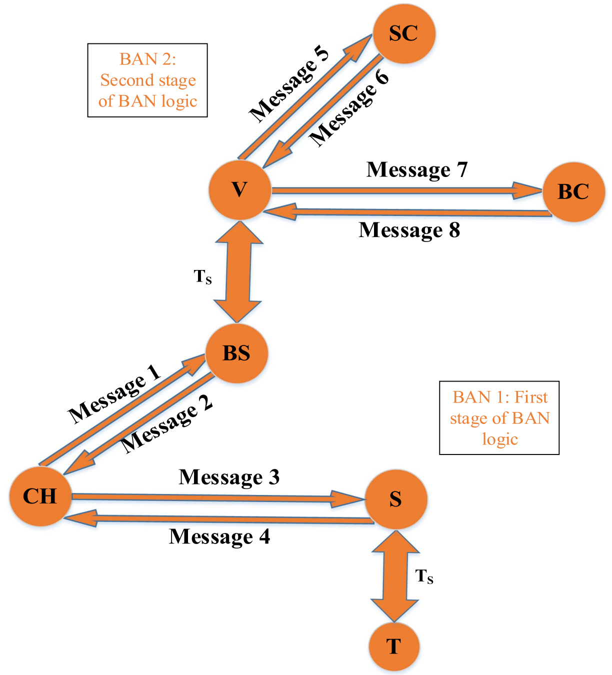


FIGURE 7. BAN logic message flow based on Kerberos protocol.

6) COMMUNICATED MESSAGES

These are defined by the Kerberos protocol [39] based on the shared-key Needham-Schroeder protocol [40]. Timestamps are considered nonce for verification of message freshness. The messages are defined as two layers of hierarchical order as shown in Fig. 7. Firstly, a timestamp T_S is defined between the sensor node (S) and target (T). The clock is considered fully synchronized between different units of the system for maintaining and confirming the freshness of data. The first layer of the message considers the communication between the base station (BS), the cluster head (CH) and the sensor node (S). CH and S are acting as principals and BS is acting as a server. The server defines the keys between principals CH and S. Whenever intending to establish a communication link with S, the cluster head CH initiates a communication request to the BS using Message 1. CH shares the node IDs of CH and S with BS as parameters. In Message 2, the BS shares the timestamp T_s and the length of session L (lifetime of the session depends on the change in RSSI value in our case). Then, in Message 3, CH shares the message with sensor S along with T_s, K_{CHS} and encrypted message shared by BS for S. Finally, in Message 4, S verifies the freshness of CH with the timestamp and received key. The messages of BAN 1 are defined below:

$$Message\ 1.\ CH \rightarrow BS : CH,\ S.$$

$$Message\ 2.\ BS \rightarrow CH : \{T_S,\ L,\ K_{CHS},\ S,$$
$$\{T_S,\ L,\ CH,\ K_{CHS}\}_{K_{SB}},\ CH\}_{K_{CS}}.$$

$$Message\ 3.\ CH \rightarrow S : \{T_S,\ L,\ K_{CS},\ CH\}_{K_{SB}}, \{S,\ T_S\}_{K_{CS}}.$$

$$Message\ 4.\ S \rightarrow CH : \{T_S + 1\}_{K_{CS}}.$$

For maintaining synchronization between two considered BAN scenarios, firstly, T_S is shared between BS and V. With the change in target position, the parameters $(RSSI,\ C_s)$ of sensor nodes changes. It results in the expiry of T_S and the base station's selection of a new set of validators. Similar to BAN 1 explained above, in the second layer of message consideration validator (V) and blockchain (BC) are the principals and the Smart contract (SC) is the server because SC selects the majority-based data (X) for storing in BC. In the second set of considered BAN logic, SC have double-fold responsibilities. Along with key establishment between V and BC, it also performs a majority-based selection of data X from the data it receives from the validator and encrypts it with key (K_{SCBC}) such that only SC and BC can read this as shown in Message 6 and 7. The messages of BAN 2 are defined below:

$$Message\ 5.\ \ V \rightarrow SC : V,\ BC.$$

$$Message\ 6.\ \ SC \rightarrow V : \{T_{SC},\ L,\ K_{VBC},\ BC,\ T_{SC},$$
$$L,\ K_{VBC},\ V,\ X\}_{K_{SCBC}}\}_{K_{VSC}}.$$

$$Message\ 7.\ \ V \rightarrow BC : \{T_{BS},\ L,\ K_{VBC},\ V,\ X\}_{K_{BCBS}},$$
$$\{V,\ T_V\}_{K_{VBS}}.$$

$$Message\ 8.\ \ BC \rightarrow V : \{T_V + 1\}_{K_{VBC}}.$$

7) IDEALIZED FORM OF THE PROPOSED SCHEME

A message in idealized form is called a formula. In an idealized form, the message is presented in encrypted form rather than cleartext form. The idealized form of the messages defined in the previous section is presented below:

$$Message\ 2.\ \ BS \rightarrow CH : \left\{T_S,\ CH \xleftrightarrow{K_{CS}} S\right\},$$
$$\left\{T_S,\ CH \xleftrightarrow{K_{CS}} S\right\}_{K_{SB}}\ \}_{K_{CB}}.$$

$$Message\ 3.\ \ CH \rightarrow S : \left\{T_S,\ CH \xleftrightarrow{K_{CS}} S\right\}_{K_{SB}},$$
$$\left\{T_S,\ CH \xleftrightarrow{K_{CS}} S\}_{K_{CS}}\right\}\ from\ CH.$$

$$Message\ 4.\ \ S \rightarrow CH : \left\{T_S,\ CH \xleftrightarrow{K_{CS}} S_{K_{CS}}\right\}\ from\ CH.$$

$$Message\ 6.\ \ SC \rightarrow V : \left\{T_{SC},\ V \xleftrightarrow{K_{VBC}} BC\right\},$$
$$\left\{T_{SC},\ V \xleftrightarrow{K_{VBC}} BC,\ X\right\}_{K_{SCBC}}.$$

$$Message\ 7.\ \ V \rightarrow BC : \left\{T_{SC},\ V \xleftrightarrow{K_{VBC}} BC\right\}_{K_{BCSC}},$$
$$\left\{T_V,\ V \xleftrightarrow{K_{CS}} BC\right\}_{K_{VBC}}\ from\ V.$$

$$Message\ 8.\ \ BC \rightarrow V : \left\{T_V,\ V \xleftrightarrow{K_{VBC}} BC\right\}_{K_{VBC}}\ from\ BC.$$

8) SECURITY ANALYSIS PROOF

This section explains the formal analysis of the proof. The idealized form of the message and considered assumption help in the proof as explained below:

CH receives Message 2 which means that

$$CH \lhd \left\{T_C,\ (CH \xleftrightarrow{K_{CS}} S)\right\},\ \left\{T_S,\ CH \xleftrightarrow{K_{CS}} S\right\}_{K_{SB}}\ \}_{K_{CB}}. \tag{5}$$

From assumptions, we have

$$CH\ |\equiv CH \xleftrightarrow{K_{CB}} BS. \tag{6}$$

Applying the message meaning rule to equations 5 and 6, we get

$$CH\ |\equiv BS\ |\sim \left\{T_{BS},\ CH \xleftrightarrow{K_{CS}} S\right\},\ \left\{T_S,\ CH \xleftrightarrow{K_{CS}} S\right\}_{K_{SB}}\ \}. \tag{7}$$

Using the break conjunction rule,

$$CH\ |\equiv BS\ |\sim \left(T_S,\ \left(CH \xleftrightarrow{K_{CS}} S\right)\right). \tag{8}$$

From assumptions, we have

$$CH\ |\equiv \#T_S. \tag{9}$$

Using nonce-verification rule

$$CH\ |\equiv BS\ |\equiv \left(T_S,\ \left(CH \xleftrightarrow{K_{CS}} S\right)\right). \tag{10}$$

Again applying break conjunction on (10),

$$CH\ |\equiv BS\ |\equiv CH \xleftrightarrow{K_{CS}} S. \tag{11}$$

Deriving (11) to the more concrete form,

$$CH\ |\equiv BS \Rightarrow CH \xleftrightarrow{K_{CS}} S. \tag{12}$$

From the jurisdiction rule, we derive the following

$$CH\ |\equiv CH \xleftrightarrow{K_{CS}} S. \tag{13}$$

This is the conclusion of the analysis of Message 2. CH passes Message 3 to S along with timestamp. S can decrypt it with the knowledge of key K_{CS}. Using the same logical steps used for Message 2, the analysis conclusion of Message 3 is shown below:

$$S\ |\equiv CH \xleftrightarrow{K_{CS}} S. \tag{14}$$

Applying message meaning and nonce-verification to equations (13) and (14), we get

$$S\ |\equiv CH\ |\equiv CH \xleftrightarrow{K_{CS}} S. \tag{15}$$

Message 4 assures that S believes CH and got the latest message from CH. The final results from the above analysis are as follows:

$$CH \mid\equiv CH \overset{K_{CS}}{\longleftrightarrow} S.$$

$$S \mid\equiv CH \overset{K_{CS}}{\longleftrightarrow} S.$$

$$CH \mid\equiv S \mid\equiv CH \overset{K_{CS}}{\longleftrightarrow} S.$$

$$S \mid\equiv CH \mid\equiv CH \overset{K_{CS}}{\longleftrightarrow} S.$$

From the final results of the analysis, it is concluded that now there is a direct belief between the cluster head and sensor node. Both believe that only CH and S can see the data and that it is always fresh, as a timestamp is used to verify its freshness. This proves our defined goal 1:

$$S \mid\equiv CH \mid\equiv CH \overset{K_{CS}}{\longleftrightarrow} S.$$

Similar steps from equations (5) to (13) can be repeated and the conclusion of the analysis of Message 6 is shown below:

$$V \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC. \tag{16}$$

V passes Message 7 to BC with timestamp. BC can decrypt it with the key K_{VBC} and analysis conclusion of Message 7 is shown below:

$$BC \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC. \tag{17}$$

Using message meaning and nonce-verification to equations (16) and (17), we get

$$BC \mid\equiv V \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC. \tag{18}$$

Message 8 assures that BC believes V and got a fresh message from V. The final results from the analysis of Messages 5–8 are as below:

$$V \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC.$$

$$BC \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC.$$

$$V \mid\equiv BC \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC.$$

$$BC \mid\equiv V \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC.$$

From the final results of the analysis, it is concluded that now there is a direct belief between the validator node (V) and the blockchain (BC). Both believe that only V and BC can see the data and that it is always fresh, as a timestamp is used to verify its freshness. This proves our defined goal 2.

$$BC \mid\equiv V \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC.$$

From the conclusion of BAN 2, we have

$$BC \mid\equiv V \overset{K_{VBC}}{\longleftrightarrow} BC. \tag{19}$$

In Message 7, $X_{K_{SCBC}}$ is forwarded to BC by V. As BC have access to X with key K_{SCBC}. Therefore, we can write it as

$$BC \triangleleft X_{K_{SCBC}}. \tag{20}$$

Applying the message meaning rule to (19) and (20), we get

$$BC \mid\equiv V \mid\sim X. \tag{21}$$

According to Message 8, before starting communication, BC crosschecks the freshness of data (X) and key (K) it got from V. Therefore, BC believes that X is fresh.

$$BC \mid\equiv \#(X). \tag{22}$$

As BC knows that SC have jurisdiction over X, i.e.

$$BC \mid\equiv SC \Rightarrow X. \tag{23}$$

which means that

$$BC \mid\equiv SC \mid\equiv X. \tag{24}$$

Applying the jurisdiction rule to (23) and (24), we get

$$BC \mid\equiv X. \tag{25}$$

We proved our goals 3 and 4 in (25) and (22) respectively. It completes the mathematical analysis of BAN logic.

B. DEFENSE FAILURE ANALYSIS OF THE SYSTEM

This section presents an analysis of defense failures under different conditions. In existing blockchain systems, if more than 50% nodes are compromised, the entire blockchain is compromised. In our proposed system, a few nodes are selected as validators, the selection of which is weighted. Therefore, truer nodes are more likely to be selected, resulting in a more secure system. Our system has a very high rate of successful defense even when the attack capability exceeds 50%. The core idea of our formulation is from [9]. However, the difference here is, we proposed the confidence score-based weighted validator selection rather than the random selection proposed in [9].

- Obscured capacity o_c: This represents the number of nonvalidator nodes in the network; that is, the nodes that do not affect system security even if they are tempered by the attacker:

$$o_c = N - V. \tag{26}$$

  The higher the value of o_c, the lower the probability of defense failure.

- Residual capacity r_c: Represents the number of validator nodes that can still be compromised after the attacker has compromised V* nodes

$$r_c = C_A - V^*. \tag{27}$$

In these equations, N is the number of sensors in the network, V is the number of validator nodes selected, V^* is the lower limit of system failure and is equivalent to $V/2$, and C_A is the attack capability (i.e., the number of nodes tempered by an attacker).

Therefore, according to [9], the probability of defense failure based on the residual capacity, obscured capacity, attack capability, and the number of validator nodes as formulated below in (28):

$$f_d = \begin{cases} 0, \quad C_A < V^* \\ \sum_{i=0}^{C_A - V^*} \dfrac{\binom{V}{C_A - i}\binom{N-V}{i}}{\binom{N}{C_A}}, & (V^* < C_A < V), \\ & (r_c < o_c) \\ \sum_{i=0}^{C_A - V^*} \dfrac{\binom{V}{i}\binom{N-V}{C_A - i}}{\binom{N}{C_A}}, & (C_A > V), (r_c < o_c) \\ \sum_{i=0}^{N-V} \dfrac{\binom{V}{C_A - i}\binom{N-V}{i}}{\binom{N}{C_A}}, & (C_A < V), \\ & (r_c > o_c) \\ \sum_{i=0}^{N-C_A} \dfrac{\binom{V}{V-i}\binom{N-V}{C_A - V + i}}{\binom{N}{C_A}}, & (C_A > V), \\ & (r_c > o_c). \end{cases} \tag{28}$$

Different cases of (28) are explained as below:

- $C_A < V^*$; f_d is always equal to 0 because the attack capability is less than the lower limit of system failure. Therefore, the system is fully secure.
- If $V^* < C_A < V$ and $r_c < o_c$, imply that $C_A < V$ and $C_A > V^*$, the probability of defense failure is the sum of all distinct possible combinations of the probability that more than half of the validator nodes are compromised out of total V validator nodes.
- For $C_A > V$, $r_c < o_c$, the attack capability exceeds the total number of validator nodes, and the probability of defense failure also increases.
- $C_A < V$ and $r_c > o_c$, in this case, the attacker can always compromise more than half of the validator nodes. Therefore, f_d is always equal to 1.
- For $C_A > V$ and $r_c > o_c$, the attacker can always compromise more than half of the validator nodes as $r_c > o_c$. Therefore, f_d is always equal to 1.

VI. RESULTS AND DISCUSSION

Simulation environment: To evaluate the security of the proposed stochastic blockchain involving weighted validator selection, the Google Colaboratory and Matlab platforms were used. We considered the random distribution of sensor nodes with an omnidirectional antenna and a single target node T. The network diameter d_0 is 300 m. The simulation parameters are presented in Table 1 and notations used in paper are described in Table II:

TABLE 1. Simulation Parameters

Parameters	Values/Model
Node distribution	Uniform distribution
Frequency	600 MHz
Channel model	Log distance path model
Shadowing effect	Zero mean gaussian random variable
Path loss exponent (γ)	3
Attacker's strategy	Random attacks
Error in estimated channel (d_{err})	Uniformly distributed in [5, 8] m
Reference distance (d_0)	300
No. of IoT edge node	3
Noise floor	−96 dBm
Blockchain	Consortium
Consensus	Weighted Validator Selection
Authentication	BAN Logic
Smart Contract	Majority-based data selection

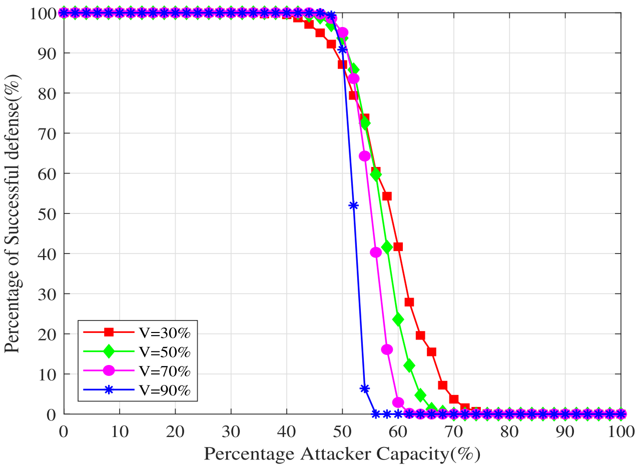


FIGURE 8. Effect of different percentage of validator on percentage successful defense with 100 nodes in the network.

A. VALIDATOR PERFORMANCE

1) EFFECT OF THE PERCENTAGE OF VALIDATOR NODES

Fig. 8 presents a comparison of the probability of successful defense (S_d) of the weighted stochastic blockchain system under varying proportions of validator nodes V and varying percentages of attack capability (C_A) for a total of 100 nodes. When the $C_A \leq 50\%$, a higher proportion of validators is contributing to a higher probability of successful defense because compromising more than 50% of the validator nodes entails compromising a greater number of nodes. However, as C_A exceeds 50%, the trend reverses. This is because when the proportion of validator nodes is excessively high (e.g., 70%

TABLE 2. **Notations**

Symbol	Description
S_dW	Probability of percentage successful defense
C_A	Attacker capacity (Number of nodes the attacker can compromise in a single attempt)
N	Total number of nodes in the network
V	Number of validator nodes selected
S_dW (Min)	Minimum value of percentage successful defense for weighted selection of validators
S_dW (Max)	Maximize value of percentage successful defense for weighted selection of validators
S_dW (Avg)	Average value of percentage successful defense for weighted selection of validators
S_dR (Min)	Minimum value of percentage successful defense for random selection of validators
S_dR (Max)	Maximum value of percentage successful defense for random selection of validators
S_dR (Avg)	Average value of percentage successful defense for random selection of validators
K_{CHS}	Key between cluster head and sensor
K_{SB}	Key between base station and sensor
K_{CS}	Key between cluster head and sensor
K_{VBC}	Key between validator and blockchain
K_{SCBC}	Key between smart contract and blockchain
K_{VSC}	Key between validator and smart contract
T_s	Timestamp between sensor and target, base station and validator

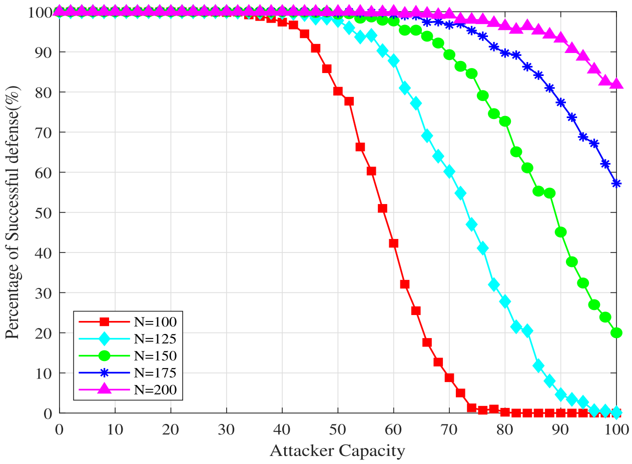


FIGURE 9. **Analysis of (S_d) for different number of nodes in network for V = 20.**

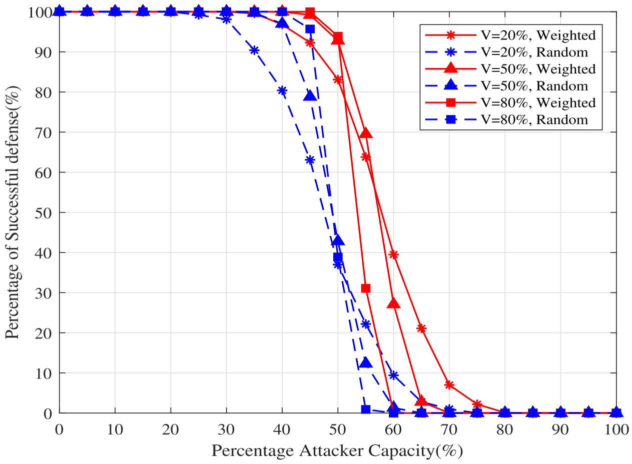


FIGURE 10. **Comparison of S_dR and S_dW for different percentage of Validator nodes with 100 nodes in the network.**

or 90%), the system begins behaving like a normal system, and the validators are easily attacked. Therefore, maintaining a lower percentage of validator nodes is recommended. When the V is 30%, even when the C_A is 65%, the likelihood of successful defense is still 20%.

2) EFFECT OF THE TOTAL NUMBER OF NODES

Fig. 9 presents a comparison of the probability of successful defense (S_d) of the weighted stochastic blockchain system for a different number of total nodes in a network with 20 validator nodes (V). Herein, C_A varies from 0 to 100. Increasing the value of N from 100 to 125 yields S_d that is approximately 40% higher for the same attack capability ($C_A = 60$). Increasing N further to 150 results in a S_d that is approximately 75% higher than it is when N = 100 and when $C_A = 70$. If N increases from 100 to 200, for the same C_A, the S_d is approximately 85%. From these results, we can conclude that for the same number of validators, as the total number of nodes increases, S_d also increases. Therefore, a system with a greater number of nodes is more secure than is one with fewer nodes.

B. PERFORMANCE COMPARISON OF THE RANDOM AND WEIGHTED SELECTION OF VALIDATOR NODES

1) EFFECT OF THE PERCENTAGE OF VALIDATOR NODES

Fig. 10 presents a comparison of the probability of successful defense with regard to the proposed weighted selection of validator nodes (S_dW) and the random selection of validator nodes (S_dR) for varying proportions of validator nodes (V) and varying percentages of attack capability (C_A) for a total of 100 nodes. The percentage of successful defense (S_d) is

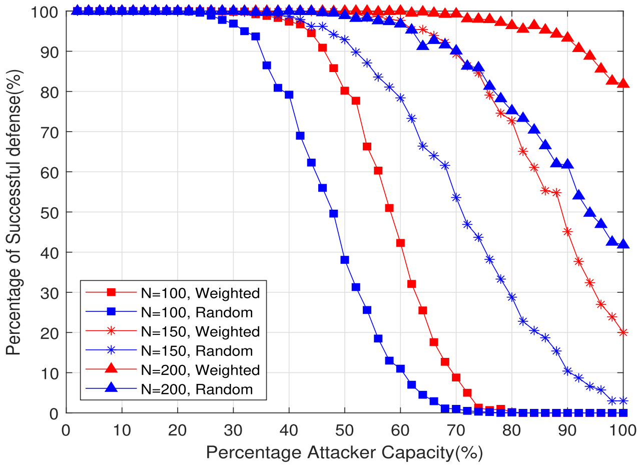


FIGURE 11. Comparative analysis of (S_d) for different number of nodes in network for V = 20 for Random and Weighted selection of validators.

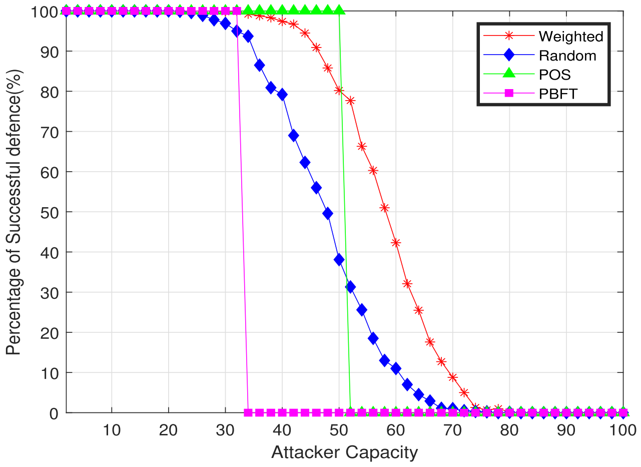
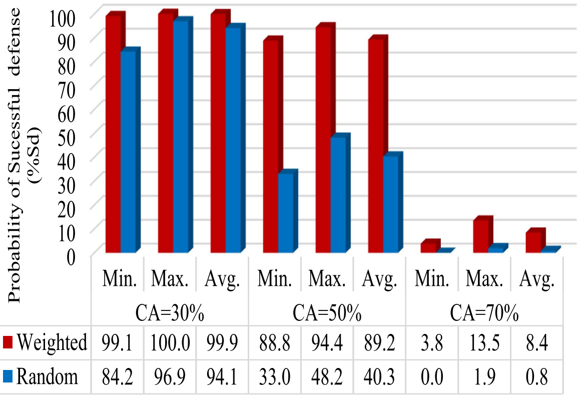


FIGURE 12. Comparative analysis of (S_d) for different consensus algorithms.

higher for the weighted validator selection than for the random validator selection. When the proportion of validators is lower, weighted validators can result in a successful defense for a higher C_A. A weighted validator scheme is more secure, providing a comparatively high S_d even when the proportion of validator nodes is high. When the $C_A = 60\%$ and the $V = 20\%$, the percentage of successful defense in weighted and random validator selection S_dW and S_dR is 40% and 10%, respectively. When the $C_A = 50\%$ and the $V = 80\%$, $S_dW = 90\%$ and $S_dR = 40\%$.

2) EFFECT OF THE TOTAL NUMBER OF NODES

Fig. 11 presents a comparison of the probability of successful defense S_d in weighted and random validator selection for varying numbers of total nodes in the network, with V = 20 and C_A varying from 0 to 100%.



	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.
	CA=30%			CA=50%			CA=70%		
🟥 Weighted	99.1	100.0	99.9	88.8	94.4	89.2	3.8	13.5	8.4
🟦 Random	84.2	96.9	94.1	33.0	48.2	40.3	0.0	1.9	0.8

FIGURE 13. Comparison table for random and weighted selection of validator nodes with different attacker capacity and (N = 100, V = 20%).

The probability of successful defense for weighted validator selection S_dW is 18% higher as compared to the S_dR for random validator selection for $C_A = 40\%$ and N = 100. As the number of nodes increases to 150, the weighted validator selection is up to 45% higher than random validator selection for $C_A = 80$. Also, for N = 200, S_dW is 25% higher than the S_dR when $C_A = 80\%$. In short, the S_dW is always higher than S_dR, and S_d increases as the number of nodes increases for the same number of validator nodes and the same attack capability.

3) PERFORMANCE COMPARISON OF DIFFERENT CONSENSUS ALGORITHMS

Fig. 12 presents the comparative analysis of S_d for a total of 100 nodes and selecting 20 validator nodes for weighted and random selection of validator nodes. For PoS and PBFT, the tolerated power of the attacker is less than 51% and 33% respectively [40]. We can observe from the graph that with the increase in attacker capacity, the S_d for PoS and PBFT falls abruptly to zero after their threshold values. However, random and weighted validator selections still have a considerable value of S_d. Further, the weighted selection of validators has a more successful defense than the random selection of validators.

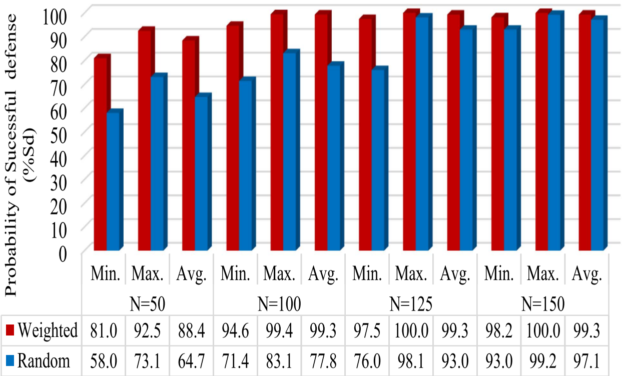
C. COMPARISON OF THE WEIGHTED AND RANDOM SELECTION OF VALIDATOR NODES FOR DIFFERENT PARAMETERS WITH REGARD TO PERCENTAGE SUCCESSFUL DEFENSE (S_d)

This section discusses the minimal, average, and maximal probability of (S_d) for different parameters. Fig. 11 presents a comparison of the S_d for the weighted and random selection of validator nodes. As the attack capability increases, (S_d) decreases. This reduction in (S_d) is greater in the random validator selection than in the weighted validator selection. When the maximal C_A is 70%, the weighted and random selection of validator nodes respectively contributes to 8% and less than 1% of (S_d).

Fig. 13 presents a comparison of the percentage successful defense for random node selection S_dR and weighted node selection S_dW under varying numbers of total nodes. As the number of total nodes increases, (S_d) also increases, whereas

TABLE 3. Comparison of Different Types of Blockchains

Property	Blockchain (BC)	Stochastic BC	Confidence score–based stochastic BC
Miner	All nodes	Random selection of validator nodes	Weighted selection of validator nodes
Miner selection	None	Random	Confidence score based weight assignment
Probability of selection as validator	Equal for all nodes	Equal for all nodes	More the value of confidence score, more the probability of being selected as validator
Average S_d (when more than 50% nodes are compromised)	0%	40%	80%



	N=50			N=100			N=125			N=150		
	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.
🟥 Weighted	81.0	92.5	88.4	94.6	99.4	99.3	97.5	100.0	99.3	98.2	100.0	99.3
🟦 Random	58.0	73.1	64.7	71.4	83.1	77.8	76.0	98.1	93.0	93.0	99.2	97.1

FIGURE 14. Comparison of the random and weighted selection of validator nodes for varying numbers of nodes (V = 20%, C_A = 40%).

the V/N and C_A/N ratios decrease. Therefore, with the increases in the total number of nodes, S_dR and S_dW also increase.

Fig. 14 presents a comparison of the S_dR and S_dW under varying numbers of nodes, with the V/N and C_A/N ratios kept constant for all values of N. Similarly, S_d increases with the increase in the number of total nodes. This suggests that even for the same proportion of validator nodes and the same attack capability, networks with a larger number of nodes have a greater likelihood of launching a successful defense than do networks with fewer nodes. The results demonstrate that the V/N and C_A/N ratios are not the only factors influencing increases in S_d; the increase in the total number of nodes also increases this probability.

D. COMPARISON OF DIFFERENT TYPES OF BLOCKCHAINS

Table 3 presents a comparison of different types of blockchains and their relevant parameters. In regular blockchains, all nodes act as validator nodes; by contrast, in stochastic blockchains, few selected nodes act as validator nodes. In simple blockchains and stochastic blockchains with random validator selection, the probability that a certain node is selected as a validator is equal. In stochastic blockchains with weighted validator selection, this probability is based on the confidence score of the node. Regarding the attack capability, when more than half of total nodes in the network are compromised, simple blockchains have a 0% probability of successful defense. By contrast, stochastic blockchains with random validator selection have a 40% probability. In stochastic blockchains with weighted validator selection, the probability is up to 80%.

VII. CONCLUSION

This article has proposed a data security mechanism for resolving data integrity and trust issues by integrating IoT with blockchain technology. The selection of a few nodes as validators mitigate problems concerning computational cost, energy, and latency. The security of data within this system is strengthened through the prudent weighted selection of validator nodes. Both excessively high and excessively low proportions of validator nodes reduce the probability of successful defense against an attack. When the number of validator nodes is low, an attack on a single validator node contributes to a high level of compromise. When validator nodes are overly numerous, they are easily located and attacked. Moreover, when the proportion of validator nodes is high, the system begins behaving like a normal system. Thus, the proportion of validator nodes must be selected very carefully to maximize data security and the probability of successful defense. A system with a larger number of nodes is always more secure than a system with a smaller number of nodes. In this study, weighted validator selection almost doubled the probability of successful defense. It has been demonstrated that a weighted stochastic blockchain is more secure than a random stochastic blockchain under the same set of parameters. With the same proportion of validator nodes and the same attack capability, stochastic blockchains with random validator selection correspond to a 40% probability of successful defense, whereas stochastic blockchains with weighted validator selection correspond to an 80% probability.

For the future work, the movable wireless IoT requires fast and efficient security and consensus mechanisms. For example, UAVs and self-driving cars are such typical application scenarios. Therefore, our proposed lightweight stochastic blockchain mechanism holds great potential for expansion into these dynamic IoT environments.

REFERENCES

[1] J. A. Manrique, J. S. Rueda-Rueda, and J. M. Portocarrero, "Contrasting Internet of Things and wireless sensor network from a conceptual overview," in *Proc. IEEE Int. Conf. Internet Things IEEE Green Comput. Commun. IEEE Cyber, Phys. Social Comput. IEEE Smart Data*, 2016, pp. 252–257.

[2] H.-N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for Internet of Things: A survey," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8076–8094, Oct. 2019.

[3] Y. Tan, J. Wang, J. Liu, and N. Kato, "Blockchain-assisted distributed and lightweight authentication service for industrial unmanned aerial vehicles," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 16928–16940, Sep. 2022.

[4] K. Boakye-Boateng, E. Kuada, E. Antwi-Boasiako, and E. Djaba, "Encryption protocol for resource-constrained devices in fog-based IoT using one-time pads," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3925–3933, Apr. 2019.

[5] D. Puthal, N. Malik, S. P. Mohanty, E. Kougianos, and C. Yang, "The blockchain as a decentralized security framework future directions," *IEEE Consum. Electron. Mag.*, vol. 7, no. 2, pp. 18–21, Mar. 2018.

[6] Y. Tan, J. Liu, and N. Kato, "Blockchain-based key management for heterogeneous flying ad hoc network," *IEEE Trans. Ind. Inform.*, vol. 17, no. 11, pp. 7629–7638, Nov. 2021.

[7] T. Alladi, V. Chamola, N. Sahu, V. Venkatesh, A. Goyal, and M. Guizani, "A comprehensive survey on the applications of blockchain for securing vehicular networks," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1212–1239, Secondquarter, 2022.

[8] Y. Liang, C. Lu, Y. Zhao, and C. Sun, "Interference-based consensus and transaction validation mechanisms for blockchain-based spectrum management," *IEEE Access*, vol. 9, pp. 90757–90766, 2021.

[9] Y. J. Chen, L.-C. Wang, and S. Wang, "Stochastic blockchain for IoT data integrity," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 373–384, Jan.–Mar. 2020.

[10] G. Lee, J. Park, W. Saad, and M. Bennis, "Performance analysis of blockchain systems with wireless mobile miners," *IEEE Netw. Lett.*, vol. 2, no. 3, pp. 111–115, Sep. 2020.

[11] D. Pavithran, K. Shaalan, J. N. Al-Karaki, and A. Gawanmeh, "Towards building a blockchain framework for IoT," *Cluster Comput.*, vol. 23, pp. 2089–2103, 2020.

[12] Z. Xiong, Y. Zhang, D. Niyato, P. Wang, and Z. Han, "When mobile blockchain meets edge computing," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 33–39, Aug. 2018.

[13] O. Onireti, L. Zhang, and M. A. Imran, "On the viable area of wireless practical Byzantine fault tolerance (PBFT) blockchain networks," in *Proc. IEEE Glob. Commun. Conf.*, 2019, pp. 1–6.

[14] L. D. Xu and W. Viriyasitavat, "Application of blockchain in collaborative Internet-of-Things services," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 6, pp. 1295–1305, Dec. 2019.

[15] W. Ni, J. Kang, C. Maple, Z. Xiong, and A. Asheralieva, "Fast and secure consortium blockchains with lightweight block verifiers," in *Proc. IEEE 3rd Int. Conf. Blockchain Comput. Appl.*, 2021, pp. 11–18.

[16] J. Kang, Z. Xiong, D. Niyato, S. Xie, and D. I. Kim, "Securing data sharing from the sky: Integrating blockchains into drones in 5G and beyond," *IEEE Netw.*, vol. 35, no. 1, pp. 78–85, Jan./Feb. 2021.

[17] M. Uddin, M. Muzammal, M. K. Hameed, I. T. Javed, B. Alamri, and N. Crespi, "CBCIoT: A consensus algorithm for blockchain-based iot applications," *Appl. Sci.*, vol. 11, no. 22, 2021, Art. no. 11011. [Online]. Available: https://www.mdpi.com/2076-3417/11/22/11011

[18] J. Ye, X. Kang, Y.-C. Liang, and S. Sun, "A trust-centric privacy-preserving blockchain for dynamic spectrum management in IoT networks," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13263–13278, Aug. 2022.

[19] H. L. J. Ting, X. Kang, T. Li, H. Wang, and C.-K. Chu, "On the trust and trust modeling for the future fully-connected digital world: A comprehensive study," *IEEE Access*, vol. 9, pp. 106743–106783, 2021.

[20] S. Lee, M. Kim, J. Lee, R.-H. Hsu, and T. Q. S. Quek, "Is blockchain suitable for data freshness? An age-of-information perspective," *IEEE Netw.*, vol. 35, no. 2, pp. 96–103, Mar./Apr. 2021.

[21] S. Lee, M. Kim, M. Kim, and J. Lee, "Timely update probability analysis of blockchain ledger in UAV-assisted data collection networks," in *Proc. IEEE Int. Conf. Commun.*, 2022, pp. 4559–4564.

[22] M. A. Ferrag and L. Shu, "The performance evaluation of blockchain-based security and privacy systems for the Internet of Things: A tutorial," *IEEE Internet Things J.*, vol. 8, no. 24, pp. 17236–17260, Dec. 2021.

[23] M. Hegde, R. R. Rao, and B. Nikhil, "DDMIA: Distributed dynamic mutual identity authentication for referrals in blockchain-based health care networks," *IEEE Access*, vol. 10, pp. 78557–78575, 2022.

[24] B. Cao et al., "When Internet of Things meets blockchain: Challenges in distributed consensus," *IEEE Netw.*, vol. 33, no. 6, pp. 133–139, Nov./Dec. 2019.

[25] T. M. Fernández-Caramés and P. Fraga-Lamas, "A review on the use of blockchain for the Internet of Things," *IEEE Access*, vol. 6, pp. 32979–33001, 2018.

[26] E. Zaghloul, T. Li, and J. Ren, "d-BAME: Distributed blockchain-based anonymous mobile electronic voting," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16585–16597, Nov. 2021.

[27] E. Androulaki et al., "Hyperledger fabric: A distributed operating system for permissioned blockchains," in *Proc. 13th EuroSys Conf.*, 2018, pp. 1–15.

[28] Z. Li, J. Kang, R. Yu, D. Ye, Q. Deng, and Y. Zhang, "Consortium blockchain for secure energy trading in industrial Internet of Things," *IEEE Trans. Ind. Inform.*, vol. 14, no. 8, pp. 3690–3700, Aug. 2018.

[29] W. Y. M. M. Thin, N. Dong, G. Bai, and J. S. Dong, "Formal analysis of a proof-of-stake blockchain," in *Proc. 23rd Int. Conf. Eng. Complex Comput. Syst.*, 2018, pp. 197–200.

[30] J. Misic, V. B. Misic, X. Chang, and H. Qushtom, "Adapting PBFT for use with blockchain-enabled IoT systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 33–48, Jan. 2021.

[31] L. Asmaa, K. A. Hatim, and M. Abdelaaziz, "Localization algorithms research in wireless sensor network based on multilateration and trilateration techniques," in *Proc. IEEE Int. Colloq. Inf. Sci. Technol.*, 2014, pp. 415–419.

[32] R. Weber, E. Mademann, O. Micnler, and S. Zeisberg, "Localization techniques for traffic applications based on robust WECOLS positioning in wireless sensor networks," in *Proc. IEEE 9th Workshop Positioning Navigation Commun.*, 2012, pp. 215–219.

[33] C. Gueguen, A. Rachedi, and M. Guizani, "Incentive scheduler algorithm for cooperation and coverage extension in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 797–808, Feb. 2013.

[34] P. S. Efraimidis and P. G. Spirakis, "Weighted random sampling with a reservoir," *Inf. Process. Lett.*, vol. 97, no. 5, pp. 181–185, 2006.

[35] V. Dedeoglu, R. Jurdak, G. D. Putra, A. Dorri, and S. S. Kanhere, "A trust architecture for blockchain in IoT," in *Proc. 16th EAI Int. Conf. Mobile Ubiquitous Syst.: Comput. Netw. Serv.*, 2019, pp. 190–199.

[36] A. Dutta and M. Chiang, "see something, say something" crowdsourced enforcement of spectrum policies," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 67–80, Jan. 2016.

[37] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Berlin, Germany: Springer Sci Business Media, 2008.

[38] M. Burrows, M. Abadi, and R. Needham, "A logic of authentication," *ACM Trans. Comput. Syst.*, vol. 8, no. 1, pp. 18–36, 1990.

[39] S. P. Miller, "Kerberos authentication and authorization system," Project Athena Technical Plan Section E. 2.1, 1988.

[40] R. M. Needham and M. D. Schroeder, "Using encryption for authentication in large networks of computers," *Commun. ACM*, vol. 21, no. 12, pp. 993–999, Dec. 1978.

SUSHILA DHAKA (Graduate Student Member, IEEE) received the M.Tech degree in electronics and communication engineering from Amity University, Gurgaon, India, in 2018. She is currently working toward the Ph.D. degree with National Yang Ming Chiao Tung University, Hsinchu, Taiwan. Her research interests include wireless data security in 6G, IoT, and blockchain and artificial intelligence.

YU-JIA CHEN (Senior Member, IEEE) received the B.S. degree and Ph.D. degree in electrical engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2010 and 2015, respectively. From 2015 to 2018, he was a Postdoctoral Research Fellow with National Chiao Tung University and Postdoctoral Research Fellow with Harvard University, Cambridge, MA, USA, from 2018 to 2019. In 2019, he was with National Central University, Taoyuan City, Taiwan, where he is currently an Assistant Professor with the Department of Communication Engineering. His research interests include low-latency communications, wireless sensing, and network security. Dr. Chen has published more than 30 articles in peer-reviewed journal and conference papers. He is holds four US patents and four ROC patents.

SWADES DE (Senior Member, IEEE) received the B.Tech. degree in radiophysics and electronics from the University of Calcutta, Kolkata, India, in 1993, the M.Tech. degree in optoelectronics and optical communication from IIT Delhi, New Delhi, India, in 1998, and the Ph.D. degree in electrical engineering from the State University of New York at Buffalo, Buffalo, NY, USA, in 2004. He is currently a Professor with the Department of Electrical Engineering, IIT Delhi. Before moving to IIT Delhi in 2007, he was a Tenure-Track Assistant Professor with the Department of ECE, New Jersey Institute of Technology, Newark, NJ, USA, from 2004 to 2007. He was an ERCIM Postdoctoral Researcher with ISTI-CNR, Pisa, Italy in 2004 and has nearly five years of industry experience in India on telecom hardware and software development, from 1993 to 1997 and 1999. His research interests include communication networks, with emphasis on performance modeling and analysis. Current directions include energy harvesting wireless networks, broadband wireless access and routing, network coexistence, smart grid networks, and IoT communications. Dr. De currently serves as the Area Editor of IEEE Communications Letters and *Elsevier Computer Communications* and an Associate Editor of IEEE Transactions on Vehicular Technology and IEEE Wireless Communications Letters.

LI-CHUN WANG (Fellow, IEEE) received the Ph.D. degree from the Georgia Institute of Technology, Atlanta, GA, USA, in 1996. From 1996 to 2000, he was with AT&T Laboratories, Atlanta, where he was a Senior Technical Staff Member with the Wireless Communications Research Department. Since 2000, he has been with the Department of Electrical and Computer Engineering, Hsinchu, Taiwan. He is currently the Chair Professor and is jointly appointed by the Department of Computer Science and Information Engineering, National Yang Ming Chiao Tung University. Dr. Wang was elected an IEEE Fellow in 2011 for his contributions to cellular architecture and radio resource management in wireless networks. He was the recipient of two Distinguished Research Awards from Taiwan's Ministry of Science and Technology (2012, 2017). He was also the co-recipients of IEEE Communications Society Asia-Pacific Board Best Award (2015), Y. Z. Hsu Scientific Paper Award (2013), and IEEE Jack Neubauer Best Paper Award (1997). His recent research interests include the areas of cross-layer optimization for wireless systems, data-driven radio resource management, software-defined heterogeneous mobile networks, Big Data analysis for industrial Internet of Things, and AI-enabled unmanned aerial vehicular networks. He holds 26 US patents, and has published more than 300 journal and conference papers, and co-edited the book, *Key Technologies for 5G Wireless Systems* (Cambridge University Press, Cambridge, U.K., 2017).