SYMBIOTIC: Reputation-based IoT Data Sharing via Blockchain

1st Ujjwal Nadhani*
Department of Computer Science
San José State University
ujjwal.nadhani@sjsu.edu

2nd Russell Tapia*
Department of Computer Science
Stanislaus State University
rtapia5@csustan.edu

3rd Praneeth Babu Marella
Department of Computer Science
George Mason University
pmarell@gmu.edu

4th Gaby G. Dagher
Department of Computer Science
Boise State University
gabydagher@boisestate.edu

***Abstract*—As the number of IoT devices grows, there will be an increased need to combine their data to represent large physical landscapes accurately. IoT devices have a diverse range of data transfer protocols, making interoperability difficult to achieve. Furthermore, most IoT data networks are permissioned, increasing data fragmentation and discouraging data sharing. In this paper, we propose SYMBIOTIC, an open, interoperable, and lightweight IoT data sharing network using a reputation-based system built on a permissionless blockchain. SYMBIOTIC addresses two major challenges in open networks: fake data and data contribution fairness by proposing a reputation-based system tracking the activity of all nodes and suggesting a data validation methodology to flag fake data. Through experimentation, we show that SYMBIOTIC can accurately correlate behavior with reputation scores, it is scalable, and robust with tolerance for up to 35% malicious nodes.**

***Index Terms*—Blockchain; IoT; Data Sharing; Reputation**

I. INTRODUCTION

Nowadays, Internet-of-Things (IoT) Devices are critical in measuring information about the physical world and relaying it virtually. By 2025, experts estimate the number of IoT devices to exceed 100 billion [1]. Therefore, it will be integral for businesses, consumers, and other IoT devices to access this vast sensor economy.

Combining data from multiple IoT devices to gain a complete perspective of a physical landscape, i.e., data fusion, is significantly hindered by fragmentation [2]. It is not easy to integrate IoT devices due to their diverse frameworks [3]. Therefore, to make data fusion more accessible, IoT devices must utilize interoperable frameworks.

Blockchain technology is both interoperable and distributed, making it the perfect backbone for IoT data fusion. There are many proposals ([4]-[20]) for using blockchains to manage both IoT devices and their data explored in the Related Works section. Most proposed systems require permission to access the network from a trusted authority.

*These authors contributed equally.

Permissioned entry exacerbates the problem of fragmentation. Gaining permission from an authority to participate in an IoT data-sharing network presents an additional barrier to collaboration. In order to fully address the problem, owners of IoT devices should be able to enter and exit the network freely. Although there are several proposals for permission-less IoT data sharing networks, our proposed system, named SYMBIOTIC, introduces a reputation system that employs a combination of data validation and network feedback.

SYMBIOTIC is designed to be an open network. The most pressing issue of open data sharing networks is parasites. Parasites are nodes that extract as much data as possible without contributing to the network. An influx of parasites eventually drives the honest contributors away from the network, as they have nothing to gain from participating. Without honest contributors, the network is functionally useless.

SYMBIOTIC is also designed to be generic enough to be used for any IoT Device data. This attribute makes detecting parasitic nodes harder. There is no standardization for the type of data being shared on the network, so proving that data is invalid becomes challenging.

SYMBIOTIC is designed to be implemented on a permission-less blockchain. Permission-less blockchains are interoperable and open, therefore preventing fragmentation. Permission-less blockchains are also designed to be tolerant of a certain amount of parasitic/malicious behavior.

The core of SYMBIOTIC is a reputation system that tracks the contributions of a Node in the network. The reputation system is the source of authority on SYMBIOTIC. In SYMBIOTIC, the decentralized reputation system is a substitute for a centralized authority. A node's reputation score also determines its level of data access in the system. Each IoT Device Owner sets a minimum reputation score requirement for their devices, ensuring only honest Owners can access their data.

SYMBIOTIC also employs an IoT data-specific machine learning model to validate data. The machine learning model detects fake data, which is critical to punishing parasitic nodes with a lower reputation score.

The contributions of this paper are as follows:

1) We propose an IoT data sharing network, SYMBIOTIC, built on a permission-less blockchain that delivers validated IoT device data to third-party requesters enabling data fusion.
2) SYMBIOTIC introduces a novel algorithm for determining the reputation of participants that incentivizes good behavior for increased data access.
3) We fully implement SYMBIOTIC on Ethereum's arrowGlacier fork to simulate behavior to show the overall effectiveness of the reputation system. Our experimental evaluation results show that SYMBIOTIC is scalable and can tolerate up to 35% malicious nodes.

II. Related Work

There are several proposals for using a blockchain or a variant of a blockchain to manage IoT devices. Similarly, there are several proposals for reputation systems built on blockchains. The first part of this section explores proposals that involve IoT device management using a blockchain ([4]-[20]). The second part of this section explores proposals for reputation systems built on blockchains([21]-[23]).

De Meo et al. [4]. proposed a distributed reputation model to track the data shared by IoT devices. The system consists of a distributed Reputation Framework(RF) and assumes that each IoT device is equipped with a trust and reputation layer(TRL). The reputation management for each IoT device is conducted on a Reputation Agent (RA), which is managed by Framework Agents(FA). The reputation system is based on feedback from other nodes. De Meo et al. [4] does not use a blockchain and is not open. Without a blockchain, the system cannot be guaranteed to be tamperproof or interoperable. Before entering the network, each IoT device must register with a Framework Agent (FA). SYMBIOTIC uses a blockchain and proposes a permission-less entry mechanism for new IoT devices.

Lueking et al. [8] propose a permission-less IoT decentralized Identity Management System using Web of Trust(WoT) stored on a distributed ledger. The WoT framework requires new nodes to receive attestations from trusted nodes. The more attestations a node has in the network, the more authority it receives. The overall framework uses the IOTA Tangle, which is not a blockchain but a distributed acyclic graph (DAG). The key feature of the IOTA Tangle is that IoT devices must verify previous transactions to include their new transaction on the block. Lueking et al. [8] use a DAG and WoT, while SYMBIOTIC uses a blockchain and a reputation-based system.

Moinet et al. [10] propose a Human-like Knowledge-based Trust model (HKT) that involves mutual surveillance by all nodes through a reputation-based system stored on a blockchain. The system relies on public key infrastructure(PKI) to resolve authentication. Devices can sign a variety of data payloads using their private keys, which are stored on the blockchain. The mutual surveillance of neighboring nodes determines the reputation system of a single node. Unlike [10], SYMBIOTIC is an open network. Furthermore, the reputation system in SYMBIOTIC relies on validator nodes instead of neighboring IoT devices.

Jiang et al. [11] and Careem and Dutta [12] establish identity and authority by tying it to an IoT device's physical location. Jiang et al. [11] propose a novel coordinate-based Byzantine consensus protocol to be used in self-driving car networks. It establishes an IoT device's Wireless Network Coordinates (WNC) by triangulating signals on the network level. Similarly, in [12], neighboring sensors verify each other and manage each other's reputation. The system of [11] and [12] require IoT devices to be physically close together with overlapping fields of detection and spend computing power validating neighboring devices. SYMBIOTIC is more generalized and does not have these requirements.

Butun and Österberg [14] and Dennis and Owen [13] provide comprehensive analyses of existing systems and define criteria for new systems. [14] specifies requirements for different types of access control policies for IoT devices in Peer to Peer(P2P) networks, using permission-ed and permission-less blockchains. Similarly, [13] specifies a generalized reputation system based on feedback and specifies various attacks on decentralized reputation systems. Dennis and Owen [13] also propose a new generalized blockchain that tracks the reputation of nodes in a peer-to-peer network. The paper suggests staking real currency with an impartial third party for low reputation nodes. In contrast, SYMBIOTIC does not require a secondary currency-based mechanism or impartial third parties. Furthermore, SYMBIOTIC is a lightweight as it is specifically designed for IoT devices.

Wu and Ansari [15] proposed an Industrial IoT (IIoT) voting mechanism that groups the IoT devices (IIoT-G). The IIoT-G can evaluate the IoTs in the network and give a high probability of correct authorization. However, they limit the application of this proposal to more resource-capable IoTs.

Novo [16] presented an architecture that uses blockchain technology to manage IoT devices, *management hub*. The management hub makes blockchain request access for the IoT devices to adapt to the limitation of IoT devices using multiple smart contracts. Zhang et al. [17] accomplished the same using smart contracts. However, both systems are designed for a permissioned network.

There are several more IoT Identity and Access Management (IAM) systems implemented using a blockchain, namely Venkatraman et al. [18], Bouras et al. [19], Nuss et al. [20], Liu et al. [21], Cash et al. [22], Bou Abdo et al. [5] and Zhu et al. [23]. [18] uses a blockchain ID management system with a smart contracts on Ethereum and designed for a closed network of businesses. [19] uses multiple consortium blockchains to propose a system with certificate authorities and channels to exchange IoT data. Similarly, [20] and [21] uses a private blockchain on Hyperledger to track IoT identities and access control policies. Cash et al. [22] proposes a tiered system consisting of protected permission-ed clusters interconnected by a permission-less network built on Ethereum. Zhu et al. [23] propose a system for classifying IoT devices based on its behavior in a smart home setting.

Papers	Network Access		Authority			
					Reputation	
	Open	Permissioned	Location	Web of Trust	Data Validation	Network Feedback
De Meo et al. [4]		✓				✓
Abdo et al. [5]		✓				✓
Liu et al. [6]		✓				✓
Debe et al. [7]		✓				✓
Lueking et al. [8]	✓			✓		
Yang et al. [9]		✓			✓	
Moinet et al. [10]		✓			✓	
Jiang et al. [11]	✓		✓		✓	
Careem & Dutta [12]	✓		✓		✓	
Dennis & Owen [13]	✓	✓				✓
This paper: SYMBIOTIC	✓				✓	✓

Table I: Comparative evaluation of main features in closely related work.

These systems were designed for closed networks with a centralized, predetermined list of approved IoT devices. With a centralized source of truth, it is trivial to determine the identity and authorization of nodes. SYMBIOTIC is an open network without a predetermined, centralized list of trusted nodes and therefore proposes a reputation-based system to build trust organically.

There are also several blockchain-based reputation systems, namely Yang et al. [9], Liu et al. [6], and Debe et al. [7]. Yang et al. [9] introduce a blockchain-based reputation system to ensure secure communications in vehicular networks. Yang et al. [9] use mutual surveillance as the backbone of its reputation system. Liu et al. [6] propose a blockchain-based reputation system that uses proof of stake and is optimized for protecting the privacy of the reviewers on retail shopping. Debe et al. [7] propose a system for tracking the reputations of IoT Fog Nodes through feedback from IoT devices, which is tracked by reputation managers. These papers require either a trusted authority or a reputation manager to oversee the network. SYMBIOTIC does not require any authorities to assign and track reputation.

Table I compares the main features of the proposals most closely related to SYMBIOTIC.

III. Problem Formulation

A. Overview

An open data sharing network is integral to fully depicting the physical environments described by a specific IoT device's data. With a healthy data sharing network, individuals will have the capacity to unlock vast amounts of data stored in devices they would not ordinarily be able to access. The primary concerns with this system include preventing malicious behaviors, enforcing fairness, and accounting for the limitations of IoT devices.

B. Malicious Behaviors in Open Networks

In an open network, there is no way of preventing malicious Owners from joining. Malicious Owners could set up IoT devices providing fake data to access legitimate device data. Simply making the network permission-ed by designating an authority to prevent malicious Owners from entering the system is a potential solution. However, permission-ed networks present a bottleneck to new Owners wishing to join the network. New Owners now have to comply with the rules and navigate the barriers to entry determined by the authority. Additionally, the authority must establish trust with the new Owners, presenting yet another challenge. Therefore, it is preferable that the system roots out malicious behavior organically while being an open network.

C. Fairness

A healthy data-sharing network must enforce fairness. Each Owner would only share their data if they were promised access to additional data in the future. Each Owner must therefore be guaranteed that the network will provide the data they want to access, assuming they contribute proportionally. Furthermore, the network should be open, allowing new Owners to join and contribute quickly and easily. More Owners will join the network if it is open, leading to more data on the network, allowing the existing Owners to access more data and incentivizing them to continue to contribute. The network must possess this positive reinforcement effect in order to be successful.

D. Limitations of IoT Devices

IoT devices are a broad classification with heterogeneous computing power and storage capabilities. Any system designed for IoT devices must have tasks that are lightweight both in computing complexity and memory.

IV. SOLUTION: SYMBIOTIC

A. Solution Overview

1) Permissionless Blockchain: The backbone of SYMBIOTIC is a permissionless blockchain because permissionless blockchains are designed to be used for open networks and mitigate malicious behavior. The consensus mechanisms in a permissionless blockchain ensure that a minority of malicious nodes cannot fake requests put onto the blockchain. The hash pointers prevent any Owner from modifying past data. Every state change in the blockchain is publicly recorded and accessible by all owners. Problem 2 (Malicious Behaviors in Open Networks) is only half-addressed by a blockchain: although malicious nodes cannot

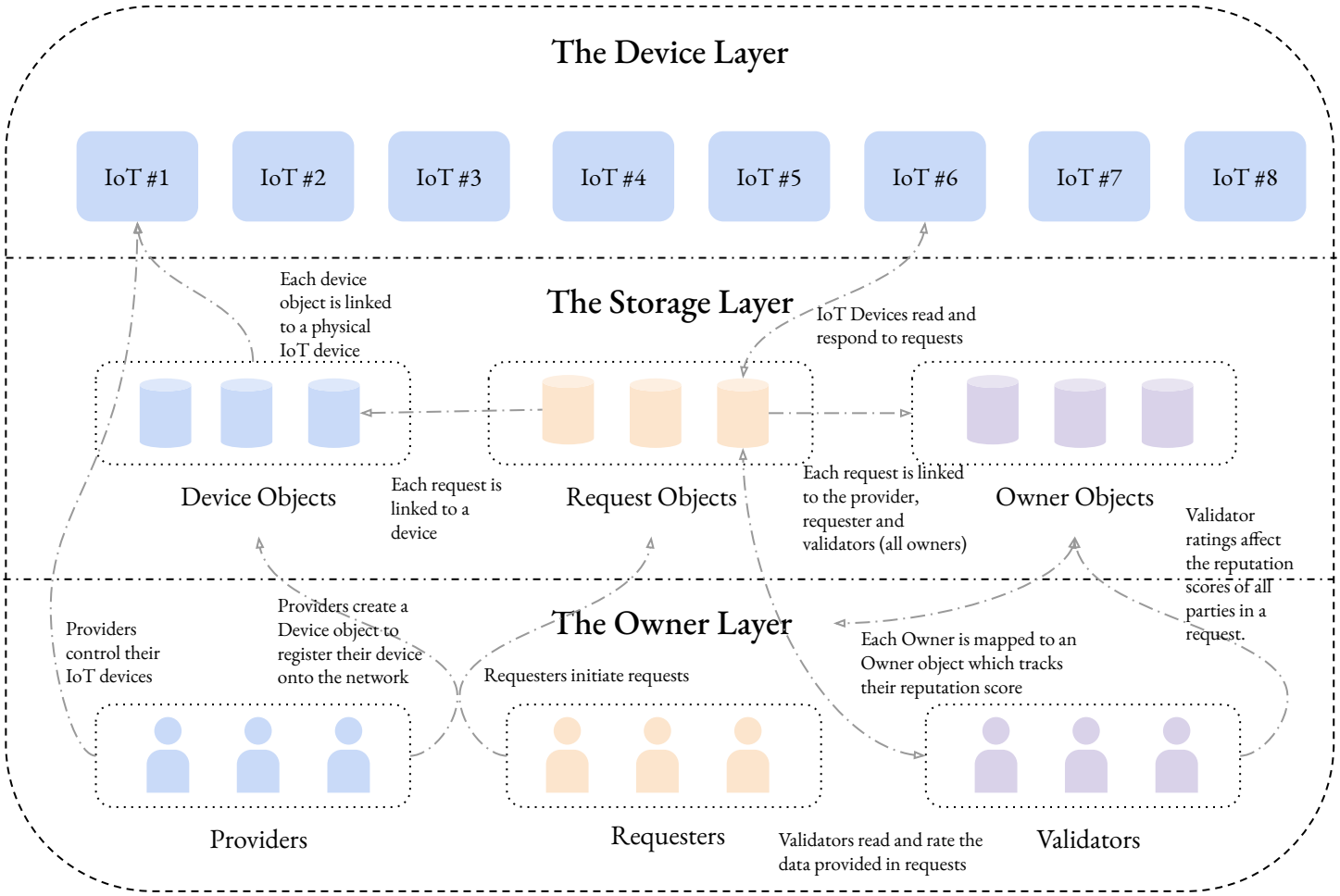

Figure 1: Overall System Diagram

fake requests, they can still respond to legitimate requests with fake data.

2) The Reputation System: In order to limit fake data on the network and ensure fairness, a reputation system is built on top of the permission-less blockchain in SYMBIOTIC. The reputation system tracks how many requests an Owner responded to, and allows independent third-party nodes known as Validators to verify the correctness of the IoT device data and maintain the blockchain. Validators use a machine learning model that determines whether data from a specific request matches previous correct data. If an Owner provides fake data, the Validators can flag it, affecting the Owner's reputation score. Requesters will be hesitant to request data from Owners with low reputation scores.

The reputation system also tracks the number of requests fulfilled to the number of requests made. If an owner acts as a parasite and requests significantly more data than they are providing, their reputation score will be lower. Owners are incentivized to keep their reputation scores high to access more data in the system. Each Owner can set a minimum reputation score requirement for a requester, denying data access to low-reputation Owners. To sum it up, the system enforces fairness and the provision of valid data through its reputation system. The reputation system dictates how much data an Owner can access, and therefore Owners are incentivized to keep their reputation scores high and act faithfully. Therefore, the reputation system in SYMBIOTIC solves Problem 3(Fairness) and combines with the blockchain to solve Problem 2(Malicious Behaviors in Open Networks).

3) Validators: To sustain a blockchain, a high amount of computational resources is required. Similarly, running a machine learning algorithm to validate data requires a lot of computational resources. IoT devices cannot perform these tasks. These tasks are outsourced to Owners to volunteer to act as Validators and dedicate computational resources to sustain the network. Block Validators maintain the blockchain by accepting new requests and verifying that the requester has a high enough reputation score to access the data they want. Before data is provided to the requester, Data Validators run a machine learning algorithm to verify that the data is authentic. The requester, which could be another Owner or even a Device, is now confident that they were given valid data without having to perform any additional computations. Validators are also rewarded with higher reputation scores as

they are integral to sustaining the network. Therefore, the Validators address the computation power aspect of Problem 4 (Limitations of IoT Devices).

B. Network Structure and Blockchain Interaction

The entire network can be subdivided into a Device layer, an Owner layer, and a Storage layer. Figure 1: Overall System Diagram illustrates all three layers and their interactions.

The IoT Devices reside in the Device layer and are assumed to be connected to the network at all times so that they can read requests, process them, and post data keys to approved requests. They are constantly reading from the Storage layer to find approved requests assigned to them. Upon reading a request, they wait for Validators to join and post the data access keys for each validator. If the Validators deem that the data provided is correct, they post the data access keys for the Requester and successfully fulfill the request.

The Owner layer is the most active part of the network. Owners connect to the network for three reasons: managing their Devices onto the network, submitting a Request for data, and earning a reputation by acting as a Validator. They dictate their device's access control policy, download the requested data, and validate the IoT device data for Requesters. Each Owner can be a Requester, Provider, or Validator in the network at any given time.

The Storage layer is implemented on a permissionless blockchain. The Storage layer is further divided into three subcomponents, Device objects, Request objects and Owner objects. Device objects store the metadata and access control policies of all devices in the network. Request objects store all addresses participating in a request, namely the Requester, the Provider, and the Validators. They also store the data access keys provided by the Device and each validator's data rating. Lastly, the Owner object stores all information necessary to calculate the reputation score of an Owner. The Validators in the Owner layer are responsible for approving all state changes in the Storage layer.

Although the system is decentralized, the entity that initializes the system must provide a pre-trained machine learning model that can be used to validate the data on the network. As the network grows, the model will improve with it as validators use it to find the correctness of more data. After the system with the model has been uploaded, the uploading entity no longer has any special access or control over the system. They will be treated just like any other node in the system.

After the system is set up, each Owner registers their device and provides a minimum reputation score as an access control mechanism. If Owners would like to restrict access to their devices, they can specify a higher minimum reputation score for requesters wanting to access their device data.

Once devices have been registered, requesters can query the blockchain to find devices. Once they find the device they are looking for, they submit a request to the network. Algorithm 1: Request Flow describes the process of handling a request. Note that each step in the algorithm represents a state change in the blockchain, and each state is approved by Block Validators.

Algorithm 1: Request Flow

Input: Request with Device D and Metadata M
Result: Validated IoT Device Data

1) A requester RQ submits a request to the network metadata M and Device D.
2) Block validators calculate the reputation score of RQ and verify that it is higher than the minimum reputation score S required to access D.
   - If (RQ's reputation score) $<$ S, the Request is rejected.
3) Owners volunteer to act as Data Validators in that Request.
4) D posts data access keys for all participating Data Validators onto the blockchain. Each data access key is encrypted with the public key of the corresponding Data Validator.
5) Each Data Validator posts their rating of Valid or Invalid to the blockchain after running the machine learning algorithm.
6) Once 5 Validators have posted their ratings, the data is determined to be valid or invalid.
   - Based on each Validator's vote and reputation score, T is calculated.
   - $V_1 ... V_5$ represent the Data Validators. $\rho(O)$ calculates the reputation score of the Data Validator. $\rho(O)$ is explored in the next section.

$$T = \sum_{i=1}^{5} \alpha(V_i)\rho(V_i)$$

$$\alpha(V) = \begin{cases} -1 & \text{if V.vote == invalid} \\ 1 & \text{if V.vote == valid} \end{cases}$$

   - If $T \geq 0$, the data is deemed to be valid.
   - Through experimentation, we found that a quorum size of 5 was enough to overrule dishonest votes.
7) If the data is considered Valid, D posts the access key for RQ, completing the Request.

At the end of the Request, the reputation score of each participant(Requester, Provider, Validators) is updated. Positive behaviors are awarded higher reputation scores which ultimately allows more access to data. Negative behaviors are punished with lower reputation scores which ultimately restricts access to data. The following section explores the specifics of the reputation calculation.

C. The Reputation System

The purpose of the Reputation System is to enforce good behavior and act as an access control mechanism for Device access. Owners are therefore incentivized to keep their reputation scores high for increased data access. This section discusses the initial formulation for the Reputation system.

This initial formulation was tested in the experimentation evaluation section to determine its effectiveness.

An Owner can choose to act as a Requester, a Provider, or a Validator. The reputation system is the sum of all the contributions made, as all three roles are weighted equally. Generally, the Reputation ρ of an Owner O can be described as

$$\rho(O) = 33r + 33p + 33v$$

where r is their Requester score, p is their Provider score, and v is their Validator score.

To standardize reputation scores, the number of participants in the system is tracked. When a new Owner performs any action in the system as a Requester, Provider, or Validator, they are counted as a participant.

The primary objective of the Requester score, r, is to ensure a proportional amount of data contribution to requests. Initially, the desired proportion of data contributions to data requests was chosen to be 1:3. This ratio can be tweaked depending on the type of data being exchanged on the network and the requirements of the participants. Therefore,

$$r = \frac{4p}{p+t}$$

where p is the provider score and t is the total number of requests made. As t, the number of requests made increases, r decreases. As p, the provider score increases, and r also increases. Therefore, this constant ensures that Owners do not act as parasites and request far more data than they provide.

The primary objective of the provider score, p, is to track an Owner's data contributions to the network relative to the average number of data contributions A_d. For every request an Owner fulfills, the Validators determine if the data is valid or invalid (see Algorithm 1). If the data is deemed valid, then the Owner is rewarded. If the data is deemed invalid, the Owner is punished for attempting to deceive a requester with fake data. The constant p is also normalized by the average number of requests fulfilled per Owner. Therefore,

$$p = \frac{\sum_{i=1}^{n} \beta(R_i)}{A_d}$$

$$\beta(R) = \begin{cases} -1 & \text{if provided data was deemed invalid} \\ 1 & \text{if provided data was deemed valid} \end{cases}$$

where $R_1 ... R_n$ represents all the requests fulfilled by the Owner O. This constant ensures that Owners do not provide fake data to deceive the network and gain access to real data.

The primary objective of the validator score, v, is to track an Owner's validator contributions to the network relative to the average validator contributions A_v. For every action a Validator does, a majority opinion and a minority opinion exist. For Block Validators, if they try to "stop a block"[1] that eventually makes it onto the blockchain, they are in the minority. For Data Validators, if their rating does not match the majority rating, they are in the minority. If the Validators are in the majority, they are rewarded. If they are in the minority, they are penalized. Therefore,

$$v = \frac{2\sum_{i=1}^{n} \epsilon(R_i)}{A_v}$$

$$\epsilon(R) = \begin{cases} -1 & \text{if Owners's vote did not match the majority} \\ 1 & \text{if Owner's vote matched the majority} \end{cases}$$

where $R_1 ... R_n$ represents all Requests where Owner O acted as a Block or Data Validator. The assumption is that malicious/parasitic Validators will constantly act in the minority. The Validators, requesters, providers, and other Validators are in check. Therefore, v is doubled as it has a higher impact on all processes of the system.

When the average number of requests fulfilled or the average number of validator contributions is less than 0.01, the network is assumed to be new with a little history. Therefore, there is insufficient data to calculate reputation scores accurately, and everyone is assumed to have the same reputation score of 100.

Overall, the reputation system is designed to assign authority to nodes in a permission-less system. As there is no designated authority to determine or enforce the ground truth, Owners have to keep each other in check as they play different roles in the system. Block Validators keep Requesters in check, who append every Request made to the blockchain, ensuring that each Owner is requesting and contributing proportionally. Providers are kept in check by Data Validators, who submit ratings about the validity of the data to the blockchain. Validators are kept in check by each other as they vote in quorums. Each Data Validator is kept in check by the majority opinion of all Data Validators assigned to the request. Similarly, each Block Validator is kept in check by the majority opinion of all Block Validators participating in the consensus mechanism of the blockchain. In a system without a centralized authority, only the majority can be trusted to determine the ground truth. Through this reputation system, SYMBIOTIC determines the ground truth to classify an Owner's behavior and correlate it with data access.

D. Data Validation Methodology

Machine learning algorithms effectively flag anomalous data, allowing SYMBIOTIC to punish dishonest nodes. Furthermore, there is precedent for machine learning in anomaly detection, as Al-Amri et al. [24] and Sign et al. [25] reviews have covered. However, for SYMBIOTIC, the following criteria must be met:

1) Supervised ML model: The main reason for using a supervised ML model is that it is impossible to get a dataset that is large enough to encompass all types of IoT data. An unsupervised ML model cannot provide enough accuracy for anomaly detection in SYMBIOTIC. As more nodes join the network, the amount of data being shared on the network will increase significantly. Supervised ML models are lightweight enough to handle this increased data without posing a significant challenge to validators.

[1]"Stopping a block" entails hindering the consensus mechanism of the permissionless blockchain used in the implementation of SYMBIOTIC.

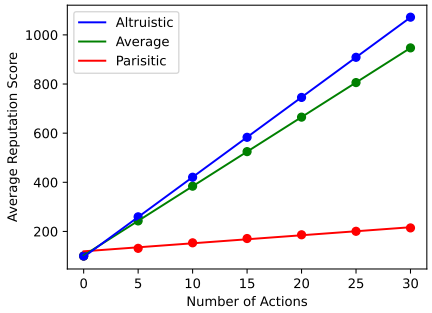

Figure 2: P = 0.15

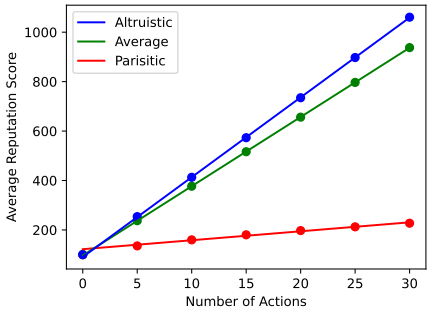

Figure 3: P = 0.25

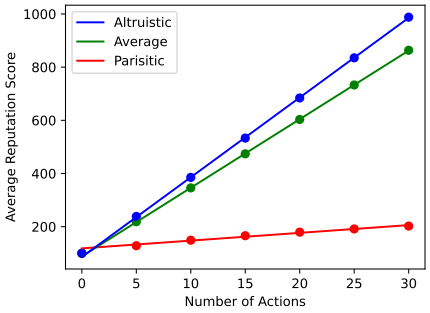

Figure 4: P = 0.35

2) Multiclass classifier: The vast diversity of IoT devices requires a multiclass classifier, where the various class categories are the different types of IoT devices. While there exist complex multiclass classifiers like tensor decomposition or convolutional neural network (CNN), we found that simpler models like Decision tree (DT), Logistic Regression (LR), and Support Vector Machine (SVM) fulfilled our purpose.

V. Experimental Evaluation

A. Experiment Setup

The primary purpose of the reputation system is to reward good behavior. The experiment's primary objective is to demonstrate a correlation between behavior and reputation score. In the experiment, there were three generalized types of behavior simulated:

1) Parisitic: A parasitic node tries to request as much data as possible without doing any work in the system. A parasitic node requests the most amount of data. It validates data incorrectly as it does not run the machine learning model. It also provides incorrect data as it has fake IoT devices to deceive the reputation system.

2) Average: An average node requests 50% less data than a malicious node. Average actually do work in the system: they run machine learning models correctly to validate data and have IoT devices that provide accurate data.

3) Altruistic: An altruistic node is an average node except that it requests 50% less data than average nodes.

1) An Owner is randomly picked from the network
2) The Owner volunteers to validate a request and waits for the Device to post its data key.
3) Devices are assumed to always be online in the network, so the Device responds with the data key as soon as the transaction is confirmed.
4) As soon as the Device responds, the Owner posts a rating. If the Owner is parasitic, the Owner posts the incorrect result from the machine learning model. If the Owner is average or altruistic, the Owner always posts the correct result from the machine learning model.
5) Steps 2 through 4 are repeated until the Owner attempts to sign up to validate five requests.
6) The Owner picks D number of Devices randomly to make their outgoing request. D is dependent on the Owner's behavior. If the Owner is malicious, D is 3. If the Owner is average, D is 2. If the Owner is altruistic, D is 1(which is 50% less).

Each simulation begins with the creation of N owners, each of whom is given exactly one device. Each owner is randomly assigned a behavior based on the simulation's P value. Owners then register themselves and their devices onto the network. After this setup process is complete, 30N actions are performed. All simulations were conducted on a local fork of Ethereum(arrowGlacier update) on a 2015 MacBook Pro with a 2.2 GHz Quad-Core Intel i7 and 16 GB of RAM.

Simulations were performed with 3 values of P: 0.15, 0.25, and 0.35 and 5 values of N: 20, 40, 60, 80, 100. There were ten folds for each distinct value of P and N. The combined reputation scores of all Ns at different P values were measured along with the runtime and the number of actions performed by each node.

B. Results and Analysis

Figures 2, 3, 4 show the reputation scores of the three different types of simulated behaviors across three different values of p. As each node performs more actions, the network can distinguish its behavior and accurately assign reputation scores. When the number of actions is 0, all nodes have the same reputation value. As nodes perform more actions, their reputation scores increase at different rates. It is evident that the behavior is correlated with the reputation scores of nodes as the number of actions increases. Across all three values of p, parasitic nodes consistently have the least amount of growth in their reputation scores. The reputation scores of average and altruistic nodes are close to each other because their behaviors are also similar. These experiments show that the overall goal of the system has been achieved.

Figures 5 and Tables II, III together show the runtime of SYMBIOTIC. Figure 5 shows that the runtime is SYMBIOTIC is linear. When there are more nodes in the system, it takes longer to perform a single action as more nodes need to reach a consensus, as shown by Figure 5. The slight amount of variability in N = 40,60 in Figure 5 was due to the simulated blockchain being in different states of consensus during the measurement of the run time calculation. Additionally, run times of applicable machine learning models were measured

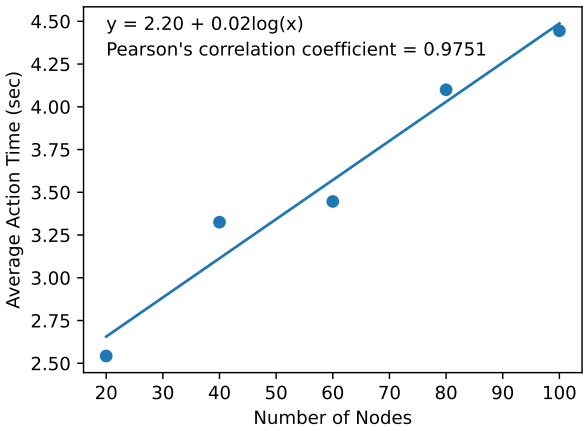

Figure 5: Average Action Runtime

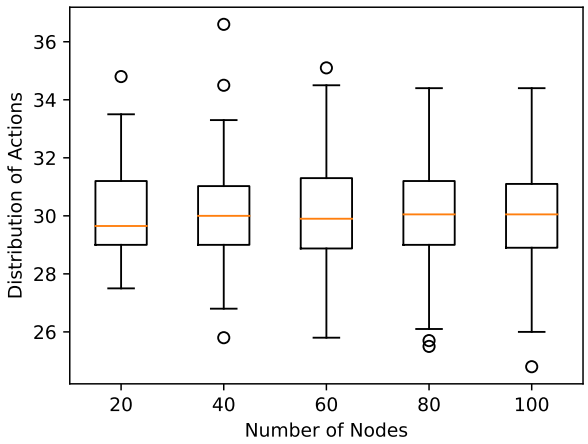

Figure 6: Action Distribution

Records	Time (sec)		
	Decision Tree	Logistic Regression	Support Vector Machine
.5 M	1.3	1.5	263
1 M	2.5	2.5	529
1.5 M	4.7	4.6	819
2 M	5.7	6.8	1253
2.5 M	8.5	10.4	1652
3 M	12.8	13.8	2077

Table II: Validation runtime of each ML model per size of records.

Records	Time (sec)		
	Decision Tree	Logistic Regression	Support Vector Machine
100 K	0.7	34	27
200 K	1.3	58	51
300 K	1.9	85	97
400 K	2.8	111	209
500 K	4.1	146	247

Table III: Fit runtime of each ML model per size of records.

using Google Colaboratory. The KDDcup99 dataset, which has 49040 observations and 42 features, was used to generate Tables II and III .The KDDcup99 dataset was copied multiple times to produce over three million records.

Figure 6 shows the participation of nodes in the simulations. The median participation is close to 30 across all values of N, showing that each Node performs close to 30 actions, as expected. In more extensive networks($N \geq 40$), the spread of actions appears to increase as there is more variance in the system. Overall, Figures 2-4 show SYMBIOTIC is effective in correlating behavior to data access, Figure 5 along with Table II and Table III show that SYMBIOTIC is scalable, and Figure 6 shows that the participation of nodes in the experiments was realistic.

VI. Conclusion and Future Work

Through this paper, we have shown that it is possible to sustain a completely permissionless IoT data sharing network even when there are up to 35% malicious nodes. This finding represents a potential solution to the primary bottleneck of IoT data sharing networks, data fragmentation. The implementation of SYMBIOTIC on Ethereum and experimentation show that SYMBIOTIC is viable in the real world. We hope that permission-less IoT data sharing systems akin to SYMBIOTIC can be used in the future to unlock the full potential of IoT devices and the sensor economy.

The success of a system like SYMBIOTIC will be highly dependent on the data validation mechanism. The network is essentially rendered useless if fake data can be propagated in the system without detection. Currently, the system relies on outside validators to run machine learning algorithms and post results. In the future, there is a possibility of executing machine learning algorithms on the blockchain. This would eliminate the need for outside output, which makes the overall network significantly more efficient and secure.

References

[1] R. Taylor, D. Baron, and D. Schmidt, "The world in 2025 - predictions for the next ten years," in *2015 10th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT)*, 2015, pp. 192–195.

[2] D. Singh, G. Tripathi, and A. J. Jara, "A survey of internet-of-things: Future vision, architecture, challenges and services," in *2014 IEEE World Forum on Internet of Things (WF-IoT)*, 2014, pp. 287–292.

[3] M. Aly, F. Khomh *et al.*, "Is fragmentation a threat to the success of the internet of things?" *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 472–487, 2019.

[4] P. De Meo, F. Messina *et al.*, "A reputation framework to share resources into iot-based environments," in *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, 2017, pp. 513–518.

[5] J. Bou Abdo, R. El Sibai, and J. Demerjian, "Permissionless proof-of-reputation-x: A hybrid reputation-based consensus algorithm for permissionless blockchains." *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 1, pp. 1 – 28, 2021.

[6] D. Liu, A. Alahmadi *et al.*, "Anonymous reputation system for iiot-enabled retail marketing atop pos blockchain," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3527–3537, 2019.

[7] M. Debe, K. Salah *et al.*, "Iot public fog nodes reputation system: A decentralized solution using ethereum blockchain," *IEEE Access*, vol. 7, pp. 178 082–178 093, 2019.

[8] M. Luecking, C. Fries *et al.*, "Decentralized identity and trust management framework for internet of things," in *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2020, pp. 1–9.

[9] Z. Yang, K. Zheng *et al.*, "A blockchain-based reputation system for data credibility assessment in vehicular networks," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–5.

[10] A. Moinet, B. Darties, and J.-L. Baril, "Blockchain based trust amp; authentication for decentralized sensor networks," 2017.

[11] Z. Jiang, Z. Cao *et al.*, "Senate: A permissionless byzantine consensus protocol in wireless networks for real-time internet-of-things applications," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6576–6588, 2020.

[12] M. A. A. Careem and A. Dutta, "Sensechain: Blockchain based reputation system for distributed spectrum enforcement," in *2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2019, pp. 1–10.

[13] R. Dennis and G. Owen, "Rep on the block: A next generation reputation system based on the blockchain," in *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, 2015, pp. 131–138.

[14] I. Butun and P. Österberg, "A review of distributed access control for blockchain systems towards securing the internet of things," *IEEE Access*, vol. 9, pp. 5428–5441, 2021.

[15] D. Wu and N. Ansari, "A trust-evaluation-enhanced blockchain-secured industrial iot system," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5510–5517, 2021.

[16] O. Novo, "Blockchain meets iot: An architecture for scalable access management in iot," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1184–1195, 2018.

[17] Y. Zhang, S. Kasahara *et al.*, "Smart contract-based access control for the internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1594–1605, 2019.

[18] S. Venkatraman and S. Parvin, "Developing an iot identity management system using blockchain," *Systems*, vol. 10, no. 2, p. 39, 2022.

[19] M. A. Bouras, Q. Lu *et al.*, "A lightweight blockchain-based iot identity management approach," *Future Internet*, vol. 13, no. 2, 2021.

[20] M. Nuss, A. Puchta, and M. Kunz, "Towards blockchain-based identity and access management for internet of things in enterprises," in *Trust, Privacy and Security in Digital Business*, S. Furnell, H. Mouratidis, and G. Pernul, Eds. Cham: Springer International Publishing, 2018, pp. 167–181.

[21] H. Liu, D. Han, and D. Li, "Fabric-iot: A blockchain-based access control system in iot," *IEEE Access*, vol. 8, pp. 18 207–18 218, 2020.

[22] M. Cash and M. Bassiouni, "Two-tier permission-ed and permission-less blockchain for secure data sharing," in *2018 IEEE International Conference on Smart Cloud (SmartCloud)*, 2018, pp. 138–144.

[23] X. Zhu, Y. Badr *et al.*, "Autonomic identity framework for the internet of things," in *2017 International Conference on Cloud and Autonomic Computing (ICCAC)*, 2017, pp. 69–79.

[24] R. Al-amri, R. K. Murugesan *et al.*, "A review of machine learning and deep learning techniques for anomaly detection in iot data," *Applied Sciences*, vol. 11, no. 12, 2021.

[25] R. Singh, N. Srivastava, and A. Kumar, "Machine learning techniques for anomaly detection in network traffic," in *2021 Sixth International Conference on Image Information Processing (ICIIP)*, vol. 6, 2021, pp. 261–266.