

# A Novel Blockchain-Reinforcement Learning Framework for Securing Wireless Sensor Networks with Energy Efficiency

Satpal Singh<sup>1</sup>, Dr. Subhash Chander<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Punjabi University spsingh.mohali@gmail.com

<sup>2</sup>Assistant Professor, Computer Science University College, Jaitu

**Abstract:** This paper introduces a novel approach to Reinforcement Learning (RL), focusing on the development and implementation of a Q-learning based algorithm. Reinforcement Learning, a critical branch of machine learning, enables agents to make decisions by interacting with their environment and learning from the consequences of their actions. Our study emphasizes the Q-learning model, a popular, model-free, off-policy algorithm that offers a robust framework for agents to learn optimal strategies in diverse settings. By iteratively updating the action-value function (Q-function) based on observed rewards and future reward estimations, our algorithm aims to achieve efficient learning and decision-making. This work contributes to the field by providing a detailed algorithmic structure, complete with mathematical formulations, that facilitates a deeper understanding of the Q-learning process and its practical applications in various domains.

## 1. Introduction

Reinforcement Learning (RL) stands at the forefront of advanced machine learning techniques, driving significant progress in artificial intelligence. It is distinguished by its unique approach, where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties [1]. This learning paradigm mimics the fundamental principles of how humans and animals learn from their experiences, making it exceptionally relevant and adaptable to a wide range of applications, from robotics to game theory.

Central to our exploration is the Q-learning algorithm, a cornerstone in the field of RL. Q-learning is particularly renowned for its model-free property, allowing agents to learn without a predefined model of the environment. This makes it incredibly versatile and applicable to situations where the dynamics of the environment are complex or unknown [2]. The off-policy nature of Q-learning further enhances its appeal, as it permits the agent to learn from actions that are outside its current policy, thereby broadening its learning scope.

Our study delves into the mathematical underpinnings of the Q-learning algorithm, presenting a detailed and structured approach to its implementation. By meticulously defining the algorithm's components, such as the state and action sets, reward function, and the Q-function, we lay a foundation for a comprehensive understanding of how Q-learning operates [3-4]. The iterative process of updating the Q-function based on immediate rewards and future reward estimates forms the crux of our discussion, highlighting the algorithm's capability to balance immediate and long-term rewards effectively.

The potential applications of this RL algorithm are vast and diverse. From optimizing decision-making processes in autonomous systems to enhancing strategies in complex games, the implications of our study are far-reaching. By providing a clear, mathematical formulation of the Q-learning algorithm, this paper aims to

contribute to the growing body of knowledge in RL and offer a resource for researchers and practitioners seeking to apply these concepts in various real-world scenarios.

## 2. Literature Review

Wireless Sensor Networks are widely used for various applications such as environmental monitoring, healthcare systems, and industrial automation. To ensure efficient data transmission and network management in Wireless Sensor Networks, routing protocols play a crucial role. One popular routing protocol for WSNs is the Low Energy Adaptive Clustering Hierarchy protocol [6]. LEACH is a clustering-based protocol that aims to optimize energy efficiency in WSNs.

LEACH works by dividing the network into clusters, with each cluster having a designated Cluster Head responsible for coordinating data transmission within the cluster [7]. The clustering approach in WSNs is crucial for optimizing energy efficiency and prolonging the lifetime of the sensor network. Low-Energy Adaptive Clustering Hierarchy is recognized as one of the most famous clustering protocols in WSNs due to its ability to reduce energy consumption. By utilizing a clustering mechanism, LEACH effectively reduces energy consumption [8-9] and extends the lifetime of the sensor network. The development of a clustering-based hierarchy protocol has significantly improved the energy efficiency in WSNs, ultimately leading to better performance in terms of system lifetime, latency, and application-perceived quality.

However, along with the benefits of WSN routing protocols, ensuring the security and integrity of data transmission in WSNs is also critical. One approach to enhancing the security of WSNs is through the use of blockchain technology [10-11]. Blockchain technology provides a decentralized and secure platform for storing and managing data in WSNs. By leveraging the decentralized and immutable nature of blockchain, the integrity and confidentiality of data in WSNs can be ensured. Blockchain technology can provide

secure and tamper-proof data storage in WSNs by eliminating the need for a central authority. One proposed solution for combining WSN routing protocols with blockchain-based security is the Reinforcement Learning Based LEACH protocol. The Reinforcement Learning Based LEACH protocol utilizes reinforcement learning techniques to enhance the security of data transmission in WSNs.

It employs a learning algorithm that adapts the routing decisions based on the network's security requirements, ensuring that data is transmitted through secure and reliable paths [12-14]. This integration of Reinforcement Learning Based LEACH protocol with blockchain technology offers a promising solution to the pressing security concerns in WSNs. By incorporating reinforcement learning, the protocol can dynamically adjust its routing decisions to adapt to changing security threats and requirements, enhancing the overall resilience of the network. This adaptability is crucial in today's dynamic and evolving threat landscape, where traditional static security measures may fall short.

Furthermore, the utilization of blockchain technology in WSNs introduces a layer of trust and transparency in data transactions. The transparency and immutability of blockchain technology help in establishing trust among the various entities involved in WSN data transactions. This trust is particularly crucial in healthcare applications, where the integrity and privacy of patient data are of utmost importance. By integrating blockchain technology into WSNs for healthcare applications, the security and privacy of patient data can be significantly enhanced, ultimately contributing to improved healthcare delivery and patient outcomes [15]. In conclusion, the combination of WSN routing protocols with blockchain-based security, particularly the Reinforcement Learning Based LEACH protocol, presents a promising approach to addressing the security challenges in WSNs.

**3. Proposed Method for Secure and Energy-Efficient WSN Using Blockchain and Q-LEACH Protocol**

Our proposed methodology focuses on securing and optimizing the energy efficiency of Wireless Sensor Networks (WSNs) through the integration of blockchain technology and the development of the Q-LEACH (Quantum LEACH) protocol. This novel approach blends Reinforcement Learning (RL) with the conventional LEACH protocol to create a more efficient and secure wireless network.

**Phase 1: Data Collection and Initialization** In the initial phase, we employ MATLAB to gather and process routing data within WSNs. This involves setting up essential parameters like the number of nodes, the number of rounds, and energy constraints, which are critical in simulating real-world scenarios. Following data collection, we integrate blockchain technology. The data processed in MATLAB is uploaded to a blockchain platform, utilizing SHA256 encryption. This step is crucial for securing the data against threats such as Sybil attacks, thereby enhancing the overall security of the WSN.

**Phase 2: Implementation of Q-LEACH Protocol** The second phase introduces the Q-LEACH protocol, an advanced iteration of the traditional LEACH. This protocol incorporates RL to facilitate smarter decision-making and path selection, optimizing energy

usage across the network. The Q-LEACH protocol involves training a model using RL techniques, where the system learns and evolves to improve path selection. This process plays a pivotal role in enhancing routing efficiency and reducing energy consumption within the network.

**Phase 3: Security Enhancement through Blockchain** To further bolster the security of the WSN, we implement Proof of Work (PoW) and Proof of Authority (PoA) algorithms within the blockchain framework. These algorithms are instrumental in protecting the network against Sybil attacks and in maintaining data integrity. The decentralized nature of blockchain enhances security, while its inherent immutability ensures the integrity of the data being managed.

**Phase 4: Simulation and Evaluation** In this phase, we simulate the WSN environment using MATLAB, deploying the Q-LEACH protocol within this controlled setting. The performance of the system is evaluated based on critical metrics such as node longevity, energy efficiency, and security against cyber threats. This simulation is vital for assessing the effectiveness of our proposed method in a realistic network environment.

**4. Hierarchical Routing Setup:**

Low energy adaptive clustering hierarchy (LEACH) is one of the main hierarchical routing protocol with aim to reduce the network energy consumption by selecting the optimal CH. The phases are as:

1. Set up Phase: In this phase nodes are initialized and the CH selection will be done using proposed trust based algorithm for cluster head selection. In this phase each node has equal chance to become CH.
2. Steady Phase: In this phase, the packets from source to destination are only sent through non malicious paths Q. Also at end of each round new CHs are formed as per Algorithm 1. All the network nodes then send their data to CH for aggregation. After data aggregation all the data is forwarded to base station BS. To avoid data collision time division multiplexing is used.

<b>Algorithm 1 for Cluster Head Selection using Trust Based Scheme</b>
Input: N – Number of nodes in cluster
Output: Elected CH
Step 1: Initialize the trust score of all the network nodes to 0.
Step 2: for each node in (N), compute Overall Trust Value as:
Direct Trust Value (DTV) as $DTV_{i,j} = \text{Final Trust Value (FTV}_j)$
Recommend Trust Value ( $RTV_{ij}$ ) = $\text{Max}(RTV)$
Overall Trust Score ( $OTS_{ij}$ ) = $DTV_{ij} + RTV_{ij}$
Step 3: do {for i : 1 to N}
Step 4: if $OTS_i < THV$ for an node
then remove the node form the cluster //Malicious node detection
else
Step 5: Elect_node (i , CH) //broadcast update in network
Step 6: for j: = 1 to $N_i$

Step 7: if $OTS_j < THV$ for an node
then remove the node form the cluster //Malicious node detection
else
Step 8: Compute M if $\{(hopcount_i) < (hopcount_j)$ and $coord\_prob_i \leq coord\_prob_j$ and $OTS_i < OTS_j \}$ then
Step 9: Elect (Role (i), member)
End if
End for
Step 10: if $i = CH$ then
Add i to cluster head set (CH)
End if
Step 11: Node with Max M, elect as CH.
Step 12: Othernode (Member CH)
<b>Algorithm 2: Secure Trust Based Routing</b>
Step 1: Initialize path discoveries for Nodes N: set $s \rightarrow d$
Step 2: for all $s \rightarrow d$ , compute Path Trust Value (PTV) as:
$PTV_{sd} = OTV_{ij}$
Step 3: if $PTV_{sd} > THS$ then
Step4: Add route to Set of valid routes (Q)
End if
End for
Step 5: for single $s \rightarrow d$ I if more than one path in Q then,
Step 6: Select path with less hop count and minimum energy usage

#### 4.1 RL-Routing Scheme

The goal of algorithm is to find the most optimal path using the proposed objective function. Initially the nodes population is initialized with feasible input required for objective functions. At each iteration the velocity and position of each node is updated as:

<b>Reinforcement Learning Algorithm</b>
Notations
<ul style="list-style-type: none"> <li>• S: Set of states in the environment.</li> <li>• A: Set of actions available to the agent.</li> <li>• R: Reward function, <math>R(s,a,s')</math>, representing the reward received after transitioning from state s to state 's' due to action a.</li> </ul>

<ul style="list-style-type: none"> <li>• <math>Q(s,a)</math>: Action-value function, estimating the value of taking action a in state s.</li> <li>• <math>\pi(a s)</math>: Policy function, representing the probability of taking action a in state s.</li> <li>• <math>\alpha</math>: Learning rate.</li> <li>• <math>\gamma</math>: Discount factor, representing the importance of future rewards.</li> </ul>
<b>Algorithm</b>
1. <b>Initialization:</b>
<ul style="list-style-type: none"> <li>• Initialize <math>Q(s,a)</math> arbitrarily for all <math>s \in S</math> and <math>a \in A</math>.</li> <li>• Choose an initial policy <math>\pi</math> (e.g., random or greedy policy based on <math>Q</math>).</li> </ul>
2. <b>For each episode:</b>
<ul style="list-style-type: none"> <li>• Initialize state s.</li> <li>• <b>For each step of the episode:</b></li> <li>• Choose action a from s using policy derived from Q (e.g., <math>\epsilon</math>-greedy).</li> <li>• Take action a, observe reward r, and next state 's'.</li> <li>• Update <math>Q(s,a)</math> using the learning rule: <math>\max(Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma \max_{a'} Q(s',a') - Q(s,a))</math></li> <li>• <math>s \leftarrow s'</math>.</li> <li>• End step loop when the terminal state is reached.</li> </ul>
3. <b>Policy Improvement:</b>
<ul style="list-style-type: none"> <li>• Improve policy <math>\pi</math> based on the updated Q values (e.g., by making it greedier with respect to Q).</li> </ul>
4. <b>Repeat:</b>
<ul style="list-style-type: none"> <li>• Repeat the process for a sufficient number of episodes or until Q converges.</li> </ul>

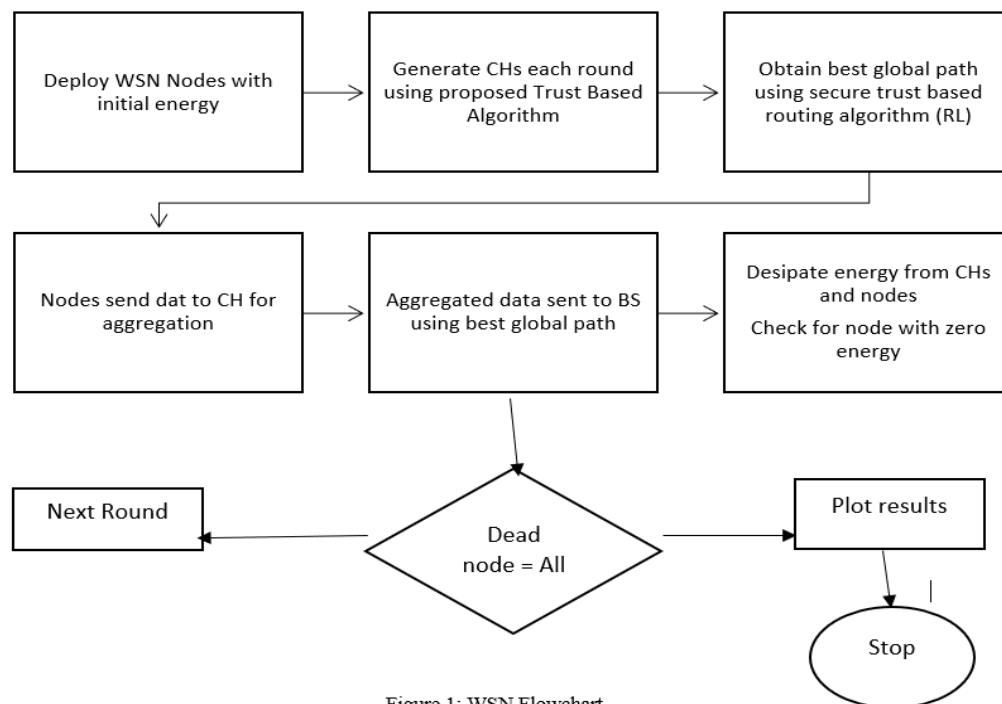


Figure 1: WSN Flowchart

### 5. Results

In this work we had integrated blockchain technology to enhance the security of data routing in Wireless Sensor Networks (WSN), leading to a more protected system. Our current objective is to augment this system's energy efficiency. To achieve this, we introduced the Q-LEACH protocol, which integrates Reinforcement Learning with the existing LEACH protocol. We analyzed Q-LEACH in comparison with three other protocols:

1. **LEACH (Low Energy Adaptive Clustering Hierarchy):** This protocol, based on hierarchical clustering, aims to minimize energy usage in WSNs. It organizes sensor nodes into clusters, electing cluster heads for data aggregation and transmission. Nodes either send data to the sink node or their cluster heads, depending on proximity and energy efficiency. The energy consumption depends on the distance to the receiver and the transmission model.
2. **TSILEACH (Two Sink Iterative LEACH):** This variant introduces two sink nodes and modifies the criteria for choosing cluster heads. A node is designated as a cluster head if it has a greater number of neighboring nodes within a certain range. It also considers multi-hop packet transmission to the sink node through several cluster

heads, allowing transmission only if the closest cluster is at a nonzero distance.

3. **Centralized LEACH:** Here, the sink node centrally manages the selection of cluster heads, focusing on the energy levels of sensor nodes. It enhances the LEACH protocol by considering the network's average energy and supporting multi-hop data transmission. Cluster heads are selected through a probabilistic model that accounts for individual node energy and the network's average energy, prioritizing nodes with above-average energy levels.

Our results demonstrate that Q-LEACH outperforms the other three protocols. In a scenario with 100 nodes and 100 rounds, the Q-LEACH protocol showed that the first and tenth nodes expired after 1200 and 1250 rounds, respectively, and all nodes expired after 1600 rounds, which is significantly better than the other protocols. When comparing total network energy over time, the other protocols exhausted their energy around 1000 seconds, while Q-LEACH sustained its energy level up to 1500 seconds. Additionally, Q-LEACH transmitted more packets to the sink node over time than the other protocols. The comparison of alive and dead nodes over time in all four protocols revealed that Q-LEACH maintained a higher number of active nodes.

<p>The Time when First Node Dies, Tenth Node Dies, and All Nodes Die</p> <table border="1"> <thead> <tr> <th>Event</th> <th>LEACH</th> <th>LEACH-C</th> <th>TSILEACH</th> <th>QLEACH</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>~900</td> <td>~1000</td> <td>~1000</td> <td>~1250</td> </tr> <tr> <td>2</td> <td>~950</td> <td>~1000</td> <td>~1000</td> <td>~1300</td> </tr> <tr> <td>3</td> <td>~1400</td> <td>~1200</td> <td>~1650</td> <td>~1700</td> </tr> </tbody> </table>	Event	LEACH	LEACH-C	TSILEACH	QLEACH	1	~900	~1000	~1000	~1250	2	~950	~1000	~1000	~1300	3	~1400	~1200	~1650	~1700	<p>Results when:</p> <p>Number of nodes- 100                  Network size- 100*100                  Number of rounds- 2500</p>
Event	LEACH	LEACH-C	TSILEACH	QLEACH																	
1	~900	~1000	~1000	~1250																	
2	~950	~1000	~1000	~1300																	
3	~1400	~1200	~1650	~1700																	
<p>The Time when First Node Dies, Tenth Node Dies, and All Nodes Die</p> <table border="1"> <thead> <tr> <th>Event</th> <th>LEACH</th> <th>LEACH-C</th> <th>TSILEACH</th> <th>QLEACH</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>~50</td> <td>~100</td> <td>~50</td> <td>~100</td> </tr> <tr> <td>2</td> <td>~50</td> <td>~150</td> <td>~150</td> <td>~200</td> </tr> <tr> <td>3</td> <td>~850</td> <td>~750</td> <td>~700</td> <td>~1050</td> </tr> </tbody> </table>	Event	LEACH	LEACH-C	TSILEACH	QLEACH	1	~50	~100	~50	~100	2	~50	~150	~150	~200	3	~850	~750	~700	~1050	<p>Results when:</p> <p>Number of nodes- 300                  Network size- 300*300                  Number of rounds- 2500</p>
Event	LEACH	LEACH-C	TSILEACH	QLEACH																	
1	~50	~100	~50	~100																	
2	~50	~150	~150	~200																	
3	~850	~750	~700	~1050																	
<p>The Time when First Node Dies, Tenth Node Dies, and All Nodes Die</p> <table border="1"> <thead> <tr> <th>Event</th> <th>LEACH</th> <th>LEACH-C</th> <th>TSILEACH</th> <th>QLEACH</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>~0</td> <td>~0</td> <td>~0</td> <td>~0</td> </tr> <tr> <td>2</td> <td>~5</td> <td>~5</td> <td>~5</td> <td>~15</td> </tr> <tr> <td>3</td> <td>~350</td> <td>~340</td> <td>~240</td> <td>~440</td> </tr> </tbody> </table>	Event	LEACH	LEACH-C	TSILEACH	QLEACH	1	~0	~0	~0	~0	2	~5	~5	~5	~15	3	~350	~340	~240	~440	<p>Results when:</p> <p>Number of nodes- 500                  Network size- 500*500                  Number of rounds- 2500</p>
Event	LEACH	LEACH-C	TSILEACH	QLEACH																	
1	~0	~0	~0	~0																	
2	~5	~5	~5	~15																	
3	~350	~340	~240	~440																	

Figure-2: Time when first node dies, tenth node dies and all nodes die

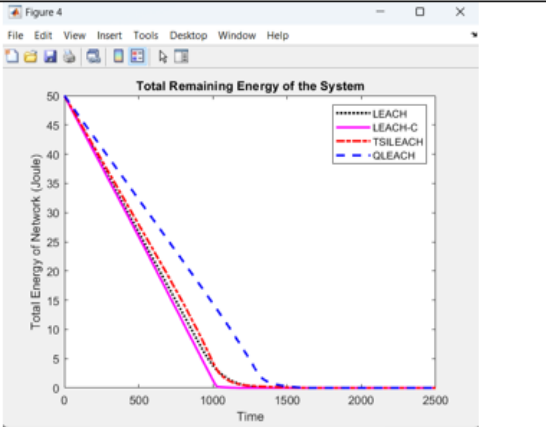
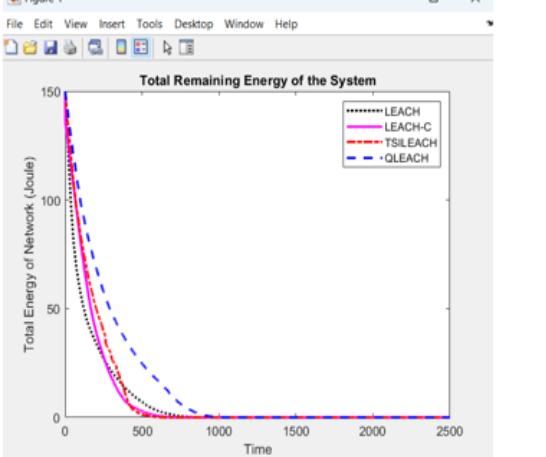
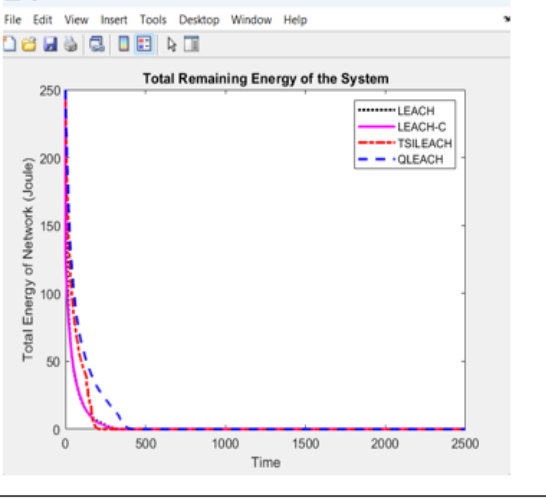
	<p>Results when:</p> <p>Number of nodes- 100 Network size- 100*100 Number of rounds- 2500</p>
	<p>Results when:</p> <p>Number of nodes- 300 Network size- 300*300 Number of rounds- 2500</p>
	<p>Results when:</p> <p>Number of nodes- 500 Network size- 500*500 Number of rounds- 2500</p>

Figure-3: Total remaining energy of the system with respect to time

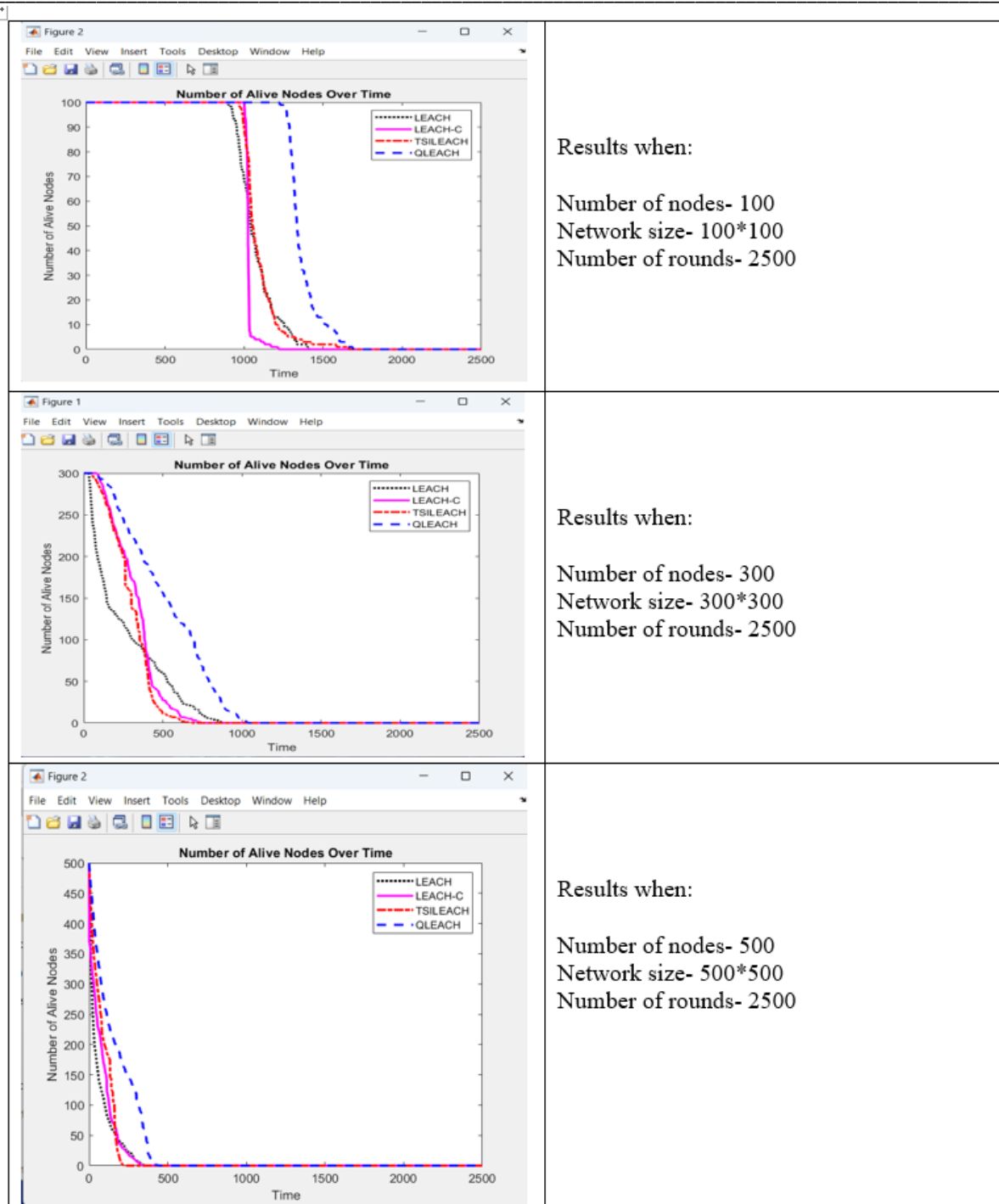


Figure-4: Number of alive nodes over time

## 6. Conclusion

The proposed research successfully develops a blockchain-based model for Wireless Sensor Networks (WSNs) to ensure enhanced security. By incorporating a Q-learning approach in MATLAB, various network parameters like the number of nodes, rounds, and energy constraints were initialized and a temporary dataset was created. This dataset, crucial for the learning process, contained significant information about the network's environment, potential transmission paths, and nodes' states and actions.

A key aspect of the research was performing a Sybil attack on the blockchain-based WSN to select an appropriate consensus mechanism for security. The study compared two mechanisms: Proof of Work (PoW) and Proof of Authority (PoA). PoW demonstrated stronger security guarantees in terms of maintaining blockchain validity, integrity, and resilience against Sybil attacks. It also offered a recovery mechanism to revert the attack. However, PoW's higher computational resource and time requirements were noted as factors to consider based on specific needs.

Post-recovery of the dataset using PoW, the research progressed to the training phase using Reinforcement Learning (RL). In this phase, sensors acted as agents to select optimal paths for data delivery, aiming to reduce transmission costs and minimize energy consumption. The culmination of this research was the development of the Q-LEACH algorithm, which enhanced the energy efficiency of the system. Q-LEACH utilized the data retrieved from the training phase and employed the Q-learning approach for better path selection, thus optimizing the network's energy efficiency. The performance of Q-LEACH was compared with LEACH, TSILEACH, and centralized LEACH, demonstrating superior results in terms of network longevity and energy management.

Overall, the study presents a significant advancement in integrating blockchain and reinforcement learning for improving the security and energy efficiency of WSNs. The research outcomes indicate a promising direction for future developments in secure and efficient network designs

## References

1. Kandris, D.; Nakas, C.; Vomvas, D.; Koulouras, G. Applications of wireless sensor networks: An up-to-date survey. *Appl. Syst. Innov.* 2020, 3, 14.
2. Yetgin, H.; Cheung, K.T.; El-Hajjar, M.; Hanzo, L.H. A survey of network lifetime maximization techniques in wireless sensor networks. *IEEE Commun. Surv. Tutor.* 2017, 19, 828–854.
3. Noel, A.B.; Abdaoui, A.; Elfouly, T.; Ahmed, M.H.; Badawy, A.; Shehata, M.S. Structural health monitoring using wireless sensor networks: A comprehensive survey. *IEEE Commun. Surv. Tutor.* 2017, 19, 1403–1423.
4. Wang, J.; Gao, Y.; Liu, W.; Sangaiah, A.K.; Kim, H.J. Energy efficient routing algorithm with mobile sink support for wireless sensor networks. *Sensors* 2019, 19, 1494.
5. Azarhava, H.; Niya, J.M. Energy efficient resource allocation in wireless energy harvesting sensor networks. *IEEE Wirel. Commun. Lett.* 2020, 9, 1000–1003.
6. Khan, Z.A.; Latif, G.; Sher, A.; Usman, I.; Ashraf, M.; Ilahi, M.; Javaid, N. Efficient routing for corona based underwater wireless sensor networks. *Computing* 2019, 101, 831–856.
7. Lee, H.C.; Ke, K.H. Monitoring of large-area IoT sensors using a LoRa wireless mesh network system: Design and evaluation. *IEEE Trans. Instrum. Meas.* 2018, 67, 2177–2187.
8. Jiang, Q.; Zeadally, S.; Ma, J.; He, D. Lightweight three-factor authentication and key agreement protocol for internet-integrated wireless sensor networks. *IEEE Access* 2017, 5, 3376–3392.
9. Shin, S.; Kwon, T. A lightweight three-factor authentication and key agreement scheme in wireless sensor networks for smart homes. *Sensors* 2019, 19, 2012.
10. Kim, T.H.; Goyat, R.; Rai, M.K.; Kumar, G.; Buchanan, W.J.; Saha, R.; Thomas, R. A novel trust evaluation process for secure localization using a decentralized blockchain in wireless sensor networks. *IEEE Access* 2019, 7, 184133–184144.
11. Guerrero-Sanchez, A.E.; Rivas-Araiza, E.A.; Gonzalez-Cordoba, J.L.; Toledano-Ayala, M.; Takacs, A. Blockchain mechanism and symmetric encryption in a wireless sensor network. *Sensors* 2020, 20, 2798.
12. Khalid, R.; Malik, M.W.; Alghamdi, T.A.; Javaid, N. A consortium blockchain based energy trading scheme for Electric Vehicles in smart cities. *J. Inf. Secur. Appl.* 2021, 63, 102998.
13. Gourisetti, S.N.; Mylrea, M.; Patangia, H. Evaluation and demonstration of blockchain applicability framework. *IEEE Trans. Eng. Manag.* 2019, 67, 1142–1156.
14. Samuel, O.; Javaid, N. GarliChain: A privacy preserving system for smart grid consumers using blockchain. *Int. J. Energy Res.* 2021, 1–17.
15. Bao, Z.; Wang, Q.; Shi, W.; Wang, L.; Lei, H.; Chen, B. When blockchain meets sgx: An overview, challenges, and open issues. *IEEE Access* 2020, 8, 170404–170420. [CrossRef]