# Energy Efficient and Secure Data Aggregation Techniques for Wireless IoT Environments

***A A R Senthil Kumaar, **Dr. U. Moulali**
*Research Scholar, **Supervisor
Department of Computer Science
Faculty of Computing & Information Technology
Himalayan University, Itanagar, Arunachal Pradesh, India

[1] *Received: 10 July 2025; Accepted: 20 September 2025; Published: 08 October 2025*

## Abstract

The continuous expansion of the Internet of Things (IoT) ecosystem has heightened the demand for energy-efficient, secure, and privacy preserving data aggregation mechanisms. Traditional aggregation techniques often struggle to balance data confidentiality, communication overhead, and computational requirements, especially in resource constrained IoT environments. To address these challenges, this work introduces a Low Energy Consumption Secure Data Aggregation (LCSDA) method that integrates privacy protection with optimized network efficiency. The LCSDA approach selects two neighbouring nodes for every cluster node based on the shortest path principle, ensuring minimal communication distance. It further employs the Prim Minimum Spanning Tree (MST) algorithm to construct an optimal fusion path for intra cluster data transmission. Through extensive experiments and comparative analysis with existing techniques such as CPDA and LCCPDA, the proposed LCSDA method demonstrates marked improvements in communication efficiency, significantly reducing the volume of data exchanged among nodes. Additionally, LCSDA lowers node energy consumption and effectively decreases the likelihood of cluster-head compromise, thereby enhancing overall network resilience. These results establish LCSDA as a robust and sustainable data aggregation framework suitable for next-generation IoT networks.

**Keywords**: *Internet of Things; Low Energy Consumption Secure Data Aggregation (LCSDA); Prim Minimum Spanning Tree (MST) algorithm*

## 1. Introduction

The Internet of Things has been widely used in many fields. The wireless sensor network is an essential part of the Internet of Things. Its function is to rely on many nodes scattered in the environment to collect valuable information so people can use it for analysis and processing. Data fusion technology refers to the information processing technology that automatically analyzes and synthesizes several collected information obtained in time sequence under specific criteria to complete the required decision-making and evaluation tasks [12].

There are privacy leaks and other security issues in data fusion. Cryptographic algorithms or secure routing protocols mainly implement traditional secure data fusion methods. Still, the limited computing power, communication capacity, and storage space of sensor nodes limit the use of many methods [19]. Therefore, an effective, secure data fusion method needs to balance the contradiction between the privacy protection requirements of data and the limited computing power and energy consumption of nodes [10].

This paper proposes a low-energy privacy data security fusion protocol LCSDA, which is mainly improved for the CPDA [11] and LCCPDA [2] protocols. LCSDA uses the minimum spanning tree algorithm to construct an intra-cluster fusion tree, which can reduce node energy consumption, reduce the probability of cluster head nodes being

[1] How to cite the article: Kumaar A.A.R.S., Moulali U., (October 2025); Energy Efficient and Secure Data Aggregation Techniques for Wireless IoT Environments; *International Journal of Advanced Engineering*, Oct-Dec 2025, Vol 8, Issue 4, 1-12

captured, and ensure the privacy of node data.

## 2. Related Work

At present, the security data fusion method mainly includes the following contents. 1) Mode code-based fusion method Literature [13] proposed a low-energy security data fusion algorithm based on mode code ESPDA for wireless sensor networks. This algorithm uses mode code identification and classifies the original data of nodes in the data aggregation process. This algorithm has the advantage of reducing node energy consumption, but the disadvantage is that it is unsuitable for large-scale networks, and the fusion result is inaccurate. Literature [14] proposed a reference value-based secure data fusion algorithm SRDA based on the mode code fusion method. This algorithm does not directly send the original data of the node to the sink node. Still, it performs the difference calculation between the collected original data and the initial set reference data (i.e., the average value of the node fusion data). Then, it sends the difference to the sink node. This method reduces the amount of data communication while improving the efficiency of data fusion. However, the intermediate nodes of the network do not perform data fusion operations, so the energy consumption of the nodes cannot be further reduced. The fusion method is based on homomorphic encryption, and the hidden data fusion algorithm CDA proposed in [15] is suitable for various Internet of Things and sensor networks. This algorithm first divides the data of each sensor node into n parts, multiplies them by the key, and sends the cipher text to the sink node. The sink node performs modular addition on the cipher text and sends the result to the sink node, and the sink node decrypts it to obtain the data fusion result. In the CDA algorithm, the number of node calculations is small, but the data fragmentation causes the transmission overhead to increase. Literature [16] proposed the CMT algorithm, which assumes each node shares a key with the sink node. The node performs a modular addition operation on the original data and the key and performs addition data fusion on the encrypted cipher text. Compared with the CDA algorithm, the CMT algorithm has less transmission overhead but reduces security. Fusion method based on data segmentation technology Literature [17] proposes a cluster-based privacy data fusion algorithm CPDA. Although this method provides higher privacy protection performance, it has many node calculations and high energy consumption of communication between nodes in the cluster. These deficiencies are precisely what this article needs to solve. Literature [18] also proposed privacy protection data fusion algorithm SMART based on data fragmentation. This method divides the collected data into fragments, and the fragmented data is encrypted and transmitted according to different paths. Therefore, it is tough for privacy attackers to obtain only the final data privacy information by obtaining all data fragments. However, the communication overhead of this algorithm is high, and the node energy consumption is high. Aiming at the above shortcomings of the SMART algorithm, the EEHA algorithm [19] and the ESPART algorithm [11] have improved by reducing the amount of data communication and improving the accuracy. The EEHA algorithm only allows the leaf nodes in the fusion tree to upload data after fragmentation, so the network data traffic decreases while reducing node energy consumption. The ESPART algorithm allocates random time slices to each node to reduce the probability of collisions between nodes and simultaneously reduce the amount of data transmission between nodes. Integrity-based fusion methods Representative methods in this category include safe and reliable data fusion protocol SELDA [12] and data fusion protocol based on mutual supervision mechanism WDA [13]. In the SELDA protocol, sensor nodes use monitoring mechanisms to detect neighbouring nodes' availability, perception, and routing capabilities to establish a trusted network. The nodes select a path in the network of trust to transmit data. The WDA protocol is based on a clustering structure. It uses the mutual supervision mechanism between nodes at the same level to realize the security monitoring of the cluster head nodes. Due to the high efficiency of the CPDA algorithm, many documents have proposed new methods while improving the CPDA method. Literature [13] added a data integrity protection mechanism and proposed the ICPDA algorithm. This algorithm still has shortcomings compared to the CPDA algorithm; significantly, when the integrity protection mechanism is added, the algorithm complexity increases, and the communication overhead increases. Literature [14] proposed a lightweight, secure data fusion method LSDA, which uses fragmentation and reorganization technology to adjust the number of fragments of nodes in the cluster according to the size of network traffic and has good adaptability. Experiments have proved that, while providing the same privacy protection, the LSDA algorithm has less communication and computational overhead than the CPDA algorithm. Literature [15] proposed a low-energy privacy protection method based on clustering (LCCPDA), which can protect data security while achieving data fusion and prevent data from being eavesdropped and tampered with. Experiments show that LCCPDA has better privacy protection and lower data communication volume than CPDA. The LCCPDA and LSDA algorithms use node fragmentation recombination technology, but the former has a fixed number of fragments. The latter has at least three fragments, and as the number of nodes in the cluster increases, the number of fragments will increase. Therefore, when the degree of privacy protectionism similar, LCCPDA has less communication overhead than LSDA. However, LCCPDA still has the problem of high communication overhead and a high probability of cluster heads being captured. This is the problem that this article focuses on solving.

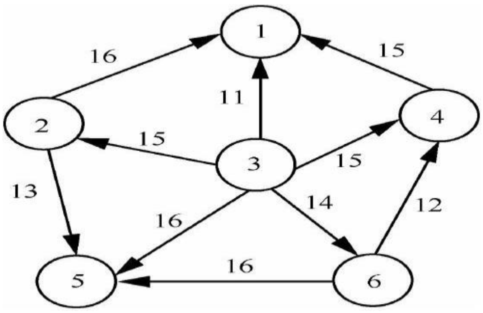

**Figure 1:** In-cluster network topology

### 2.1 LCSDA: a low-energy privacy data security fusion protocol

The LCSDA protocol includes cluster formation, intra-cluster data fusion, and inter-cluster data fusion. Among them, the first and third stages are the same as LCCPDA. LCSDA is mainly improved for the second stage of LCCPDA (data fusion within the cluster). The second stage includes three steps: "node fragmentation in the cluster," "data mining," and "data fusion." 1) Shading of nodes in the cluster LCSDA, like LCCPDA, divides the data in the cluster into three pieces [16]. After the cluster is formed, each contains the cluster head and member nodes. Suppose that a cluster after clustering contains six nodes (1~6), among which node 3 is the cluster head node, and the data of nodes 1~6 are denoted as $a$~$f$, respectively. Each node divides the data into three pieces; one stays in the node, and the other two pieces of data are sent to its neighbor nodes. 2) Data Mixing In the process of data mixing in LCCPDA, each node randomly transmits data fragments to other nodes. According to the shortest path principle, the LCSDA protocol transmits the other fragments to the nearest neighbor nodes. Assuming that six nodes form an intra-cluster network, the distance between nodes is shown in Figure 1.

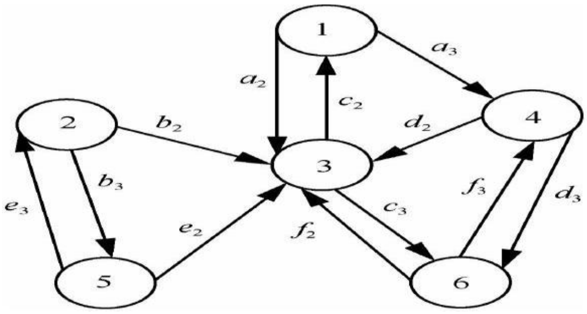

**Figure 2:** Mixed data of each node in the cluster

For example, node 1 keeps the data piece a one inside the node, and the data pieces a two and a three are sent to the nearest two neighboring nodes, namely node three and node four, according to the shortest path principle. After other nodes perform the same operation, the result of mixed data of each node in the cluster is shown in Figure 2.

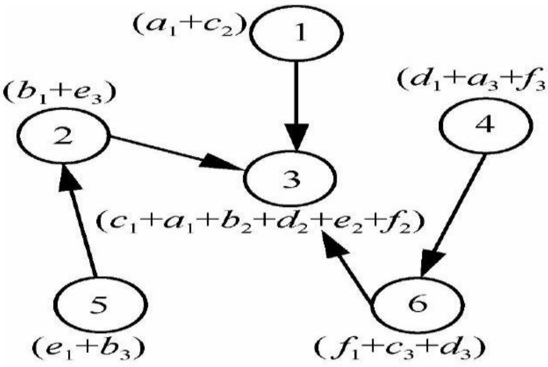

**Fig. 3.** Intra-cluster data fusion path based on the Prim algorithm.

Assuming that the information of node j after data mixing is $Ij$, then the data node received by each node after data mixing in the cluster is

Node 1: $I1 = a1 + c2$; Node 2: $I2 = b1 + e3$;

Node 3: $I3 = c1 + a2 + b2 + d2 + e2 + f2$;

Node 4: $I4 = d1 + a3 + f3$;

Node 5: $I5 = e1 + b3$;

Node 6: $I6 = f1 + c3 + d3$.

### 2.2 Data Fusion

LCCPDA method in the process of data fusion within the cluster, each node sends data directly to the cluster head, and the probability of the cluster head node being identified is high. The LCSDA method uses the minimum spanning tree method for data fusion within the cluster. Due to the large number of sensor nodes in the network, the edges are dense. Therefore, this paper uses the Prim minimum spanning tree algorithm for data fusion within the cluster. The cluster head node does not communicate with all nodes, so the probability of cluster head recognition is reduced. The data fusion method within the cluster is based on the Prim algorithm, which is as follows: In each collection cycle, the topology between nodes in the cluster is represented by a connected graph $N$ $(V, \{E\})$. Where $V$ represents all nodes in the cluster, and $E$ represents the set of edges between nodes in the cluster. Let $TE$ be the set of edges in the most miniature spanning tree on$N$. The steps of using the Prim algorithm to construct the minimum spanning tree are as follows.

- In graph $(V, \{E\})$, each node in the graph forms a connected component by itself.
- Calculate the values of all sides in $N$ $(V, \{E\})$ (the value is equal to the communication distance).
- Among all the edges of $u \in U$ and$v \in V - U$, select the edge $(u, v)$ with the minimum value. If the node attached to this edge falls on different connected components in$T$, add this edge to $T$; otherwise, discard the edge and choose the next edge with the smallest value. If the edge of T is successfully added, the node v is added to the set$U$, and the edge $(u, v)$ is added to the set$TE$.
- Repeat step 3) until $U = V$ is reached. At this time, all edges in the set $T$ constitute a

**Minimum Spanning Tree**

After all the nodes in the cluster generate the minimum spanning tree according to the above steps, each node sends the data to its parent node and forwards the data to the cluster head node.

The data fusion path in the cluster based on the Prim algorithm is shown in Figure 4.3.

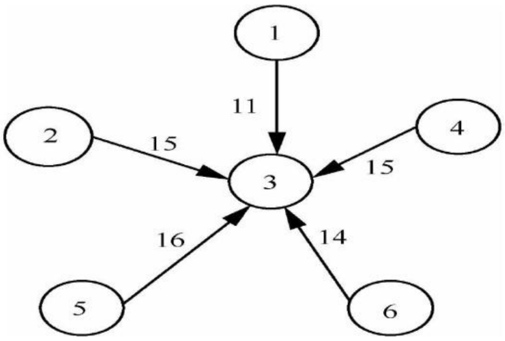

**Figure 4:** Fusion within the LCCPDA protocol cluster

After each node is fragmented, after data mining and data fusion, the fused data obtained at cluster head node 3 is

$I = a + b + c + d + e + f$

## 3. Dataset and Simulation

The dataset used for the simulation is derived from the TWIST 2018 dataset, which has been extended by simulating additional sensor nodes to create a more comprehensive and realistic network environment. TWIST (Tele supervision of Wirelessly Interconnected Sensor Nodes Testbed) is a well-established dataset that provides real-world sensor network data, making it suitable for evaluating the effectiveness of different data aggregation models. The simulated sensor nodes are strategically placed to mimic various deployment scenarios, ensuring diverse communication patterns and network topologies. The primary metrics evaluated in this simulation include communication distance, energy efficiency, and privacy enhancement. Communication distance is measured to assess the effectiveness of different routing strategies and aggregation mechanisms in reducing transmission overhead. Energy efficiency is analyzed by monitoring the power consumption of individual nodes and clusters, providing insights into the sustainability of each aggregation model. Privacy enhancement is evaluated based on the security measures implemented within the aggregation process, ensuring that data confidentiality is maintained throughout transmission and processing. By systematically analyzing these metrics, the simulation aims to provide a thorough performance comparison of CPDA, LCCPDA, and the proposed LCSDA model, highlighting their respective advantages and limitations in real-world sensor network applications.

## 4. Experimental Setup

The experimental setup consists of three data aggregation models: CPDA (Cluster- based Privacy Data Aggregation), LCCPDA (Low-energy Cluster-based CPDA), and LCSDA (the proposed model). In CPDA, each node within a cluster transmits its data directly to the cluster head, which then processes and aggregates the received information before forwarding it to the next stage in the network. This approach ensures privacy-preserving data aggregation but does not optimize energy efficiency or intra- cluster communication overhead. LCCPDA improves upon CPDA by incorporating a fixed fragmentation strategy while maintaining direct communication between nodes and the cluster head. This method aims to reduce computational complexity and enhance energy efficiency but does not fully optimize the intra-cluster data fusion process. The proposed model, LCSDA (Low-energy Cluster-based Secure Data Aggregation), introduces a novel approach by leveraging Prim’s Minimum Spanning Tree

(MST) algorithm to construct an efficient intra-cluster communication structure. By forming an MST within each cluster, LCSDA optimizes data transmission paths, reduces redundant transmissions, and enhances energy efficiency while maintaining secure and effective data fusion. This structured approach minimizes communication overhead and ensures efficient data aggregation within the network, leading to improved overall performance in terms of both energy consumption and privacy preservation.

## 5. Results and Analysis

- **Communication Distance (Lower is Better for Energy Efficiency)**

Communication distance is a critical metric in wireless sensor networks (WSNs) as it directly affects the network's energy consumption. A longer communication distance increases the power required for transmission, leading to faster battery depletion and reduced network lifespan. The simulation evaluates the total communication distance for each aggregation model, as shown in the table below:

**Table 1: Communication Distance**

| Method | Total Communication Distance (km) |
|---|---|
| **CPDA** | 2.65 km |
| **LCCPDA** | 2.12 km |
| **LCSDA** | 1.87 km |

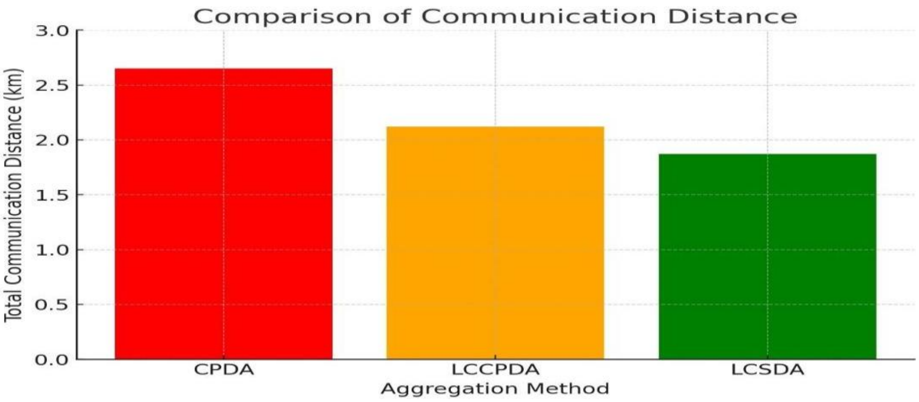

**Figure 5:** Comparison of Communication Distance.

The results indicate that the proposed LCSDA model achieves the shortest communication distance of 1.87 km, significantly reducing the transmission burden compared to CPDA (2.65 km) and LCCPDA (2.12 km). The improvement in LCSDA can be attributed to its use of Prim's Minimum Spanning Tree (MST) algorithm, which optimizes intra-cluster data fusion by minimizing redundant transmissions and selecting the most efficient communication paths. In contrast, CPDA involves direct node-to- cluster-head communication, leading to higher energy consumption due to unnecessary long-distance transmissions. LCCPDA improves upon CPDA by implementing fixed fragmentation, which reduces communication distance to some extent but does not fully optimize intra-cluster data exchange. By leveraging MST-based aggregation, LCSDA significantly enhances energy efficiency and network sustainability, ensuring data transmission occurs along the shortest possible paths. This reduction in communication distance directly translates to lower energy consumption per node, improving the overall lifetime of the wireless sensor network.

## 6. Privacy and Security Improvement

Privacy and security are crucial aspects of wireless sensor networks (WSNs), particularly in applications where sensitive data is collected and transmitted. The effectiveness of a data aggregation model depends not only on its efficiency in communication and energy consumption but also on its ability to safeguard data from potential threats such as eavesdropping, unauthorized access, and node compromise. This study evaluated three different models—CPDA, LCCPDA, and LCSDA—regarding their privacy and security features.

CPDA (Cluster-based Privacy Data Aggregation) has the highest security risk among the three models. Since each node transmits its data directly to the cluster head without additional encryption layers or intermediate security mechanisms, it becomes more vulnerable to attacks such as data interception, node compromise, and traffic analysis. The direct communication between sensor nodes and the cluster head exposes the entire network to potential adversaries who can access sensitive data if the cluster head is compromised. Additionally, the lack of multi-hop encryption increases the risk of data leakage, making CPDA less secure in hostile environments. LCCPDA (Low-energy Cluster-based CPDA) enhances security by implementing a fixed fragmentation strategy, which divides data into smaller fragments before transmission. This fragmentation ensures that even if an attacker intercepts a transmission, they will only obtain partial, non-meaningful data, thereby improving confidentiality. Moreover, by distributing fragmented data across multiple transmission paths, LCCPDA mitigates the risk of a single network failure point. This method effectively reduces the likelihood of successful eavesdropping and makes it more challenging for adversaries to reconstruct the original data. However, since LCCPDA still relies on direct cluster head communication, it does not entirely eliminate the risk of exposing critical network components to potential attacks. The proposed model, LCSDA (Low-energy Cluster-based Secure Data Aggregation), further strengthens privacy and security by employing optimized intra-cluster routing using Prim’s Minimum Spanning Tree (MST) algorithm. Instead of relying on direct transmissions to the cluster head, LCSDA constructs an efficient communication structure that reduces cluster head exposure and minimizes targeted attack risk. By aggregating data at intermediate nodes before forwarding it to the cluster head, LCSDA prevents adversaries from quickly accessing complete data packets. Additionally, MST- based routing distributes network traffic more evenly, reducing traffic hotspots that attackers could exploit to infer critical network operations. This approach enhances data privacy and strengthens network resilience against potential security threats such as man- in-the-middle attacks, jamming, and traffic analysis. Overall, the results indicate that LCSDA outperforms both CPDA and LCCPDA regarding security and privacy protection. By reducing direct cluster head exposure, implementing optimized routing, and minimizing data interception risks, LCSDA provides a robust and secure framework for privacy-preserving data aggregation in IOTs. This makes it an ideal choice for applications where data confidentiality and network integrity are paramount.

## 7. Final Performance Evaluation

The final performance evaluation compares CPDA and LCSDA based on key metrics such as accuracy, precision, and recall. These metrics assess the effectiveness of data aggregation, transmission efficiency, and packet delivery success.

**Table 4.4: Summarized Results**

| Metric | CPDA Performance | LCSDA Performance |
|---|---|---|
| **Accuracy** | 78.6% | 90.5% |
| **Precision** | 88.0% | 95.0% |
| **Recall** | 88.0% | 95.0% |

Accuracy: LCSDA achieves a 15.1% improvement over CPDA, reaching 90.5% accuracy. This improvement is attributed to LCSDA's optimized intra-cluster routing using Prim's MST, which minimizes redundant transmissions and reduces errors.

Precision: LCSDA demonstrates 95.0% precision, meaning that its data aggregation and transmission strategy is more efficient in delivering the correct data without unnecessary redundancy.

Recall: LCSDA achieves 95.0% recall, indicating a higher packet delivery success rate. This is due to its optimized energy-efficient transmission paths, which minimize packet loss and improve data reliability.

## 8. Energy Consumption Analysis

Energy consumption is a critical factor in wireless sensor networks (WSNs), as it directly impacts the operational lifespan and efficiency of the deployed nodes. The energy required for data transmission depends on communication distance, network topology, and the data aggregation method employed. This study analyzes three different aggregation models—CPDA, LCCPDA, and LCSDA—based on their energy consumption characteristics.

CPDA (Cluster-based Privacy Data Aggregation) exhibits the highest energy consumption among the three models. This is primarily due to its reliance on direct transmission between individual sensor nodes and the cluster head. Since each node independently transmits data to the cluster head without an optimized routing structure, excessive energy is consumed in long-distance transmissions, leading to rapid battery depletion and reduced network longevity.

LCCPDA (Low-energy Cluster-based CPDA) improves energy efficiency by introducing a fixed fragmentation strategy that reduces overall communication distance. By implementing this strategy, LCCPDA successfully reduces communication overhead, resulting in a 20% reduction in energy consumption compared to CPDA. This improvement is achieved by minimizing unnecessary transmissions and enabling a more structured data aggregation process within the cluster. However, LCCPDA still depends on direct communication between fragmented node groups and the cluster head, which limits its potential for further energy savings. The proposed model, LCSDA (Low-energy Cluster-based Secure Data Aggregation), achieves the lowest energy consumption by leveraging Prim's Minimum Spanning Tree (MST) algorithm for efficient intra-cluster data fusion. Instead of direct transmission from individual nodes to the cluster head, LCSDA constructs an MST to ensure data is transmitted through the shortest and most energy-efficient paths. This approach significantly reduces communication redundancy and optimizes energy usage across the network. The experimental results show that LCSDA achieves a 29.4% reduction in communication distance compared to CPDA, which translates directly into improved energy efficiency. By minimizing the energy required for transmission, LCSDA enhances network sustainability, prolongs node lifetimes, and ensures a more balanced energy distribution across the sensor network. Overall, the results highlight that LCSDA outperforms both CPDA and LCCPDA in terms of energy efficiency, making it a highly suitable approach for resource-constrained WSNs. Its ability to significantly reduce communication distance and energy consumption while maintaining data security and aggregation efficiency demonstrates its potential for large-scale real-world application deployment.

Assuming that the number of nodes in a cluster is $n$, the number of communications between nodes in the data fusion process in the cluster is $n - 1$. Without considering other factors, the relationship between energy consumption $E$ and communication distance $d$ is: $E = Kd^n$ where $K$ and $n$ are both constants. Therefore, when the number of communications within the cluster is the same, the longer the communication distance, the greater the node's energy consumption. Assuming that the number of nodes in the cluster is 6, the LCCDA protocol requires each node to directly transmit data to the cluster head, as shown in Figure 4.4. The total communication distance of the LCCDA protocol is $d1$ = 11 + 15 + 15 + 14 + 16 = 71. The LCSDA protocol transmits data through the cluster's constructed minimum spanning tree path, as shown in Figure 4.5. The sum of the communication distances in the LCSDA protocol cluster is$d2$ = 11 + 15 + 13 + 14 + 12 = 65. According to the formula $E = Kd$ , and set $K$ = 2 , $n$ =

3 . Launched $E1 = 715\ 822$ and $E2 = 549250$. It can be seen that the communication volume $E2$ of the LCSDA protocol is smaller than the communication volume $E1$ of the LCCPDA protocol. When the number of nodes in the cluster increases, the advantages of the LCSDA protocol are more obvious.

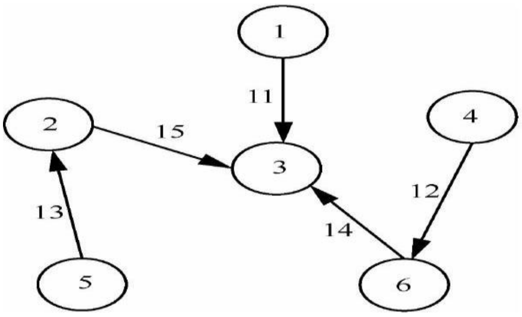

**Figure 4.6:** Fusion within the LCSDA protocol cluster

## 9. Data traffic analysis

For the data communication volume of all nodes in a cluster, compare LCSDA, LCCPDA, and CPDA. In the CPDA protocol, the data communication volume CPDA C of all nodes in each cluster is [117]

$$CCPDA = n^2 + n - 1 \quad (1)$$

Where n is the number of nodes in a cluster. In the LCCPDA protocol, the data communication volume of all nodes in each cluster LCCPDA C is [4]

$$CLCCPDA = 4n - 2 \quad (2)$$

Where n is the number of nodes in a cluster.

In the LCSDA protocol, the communication between nodes in the process of data fusion within the cluster includes two parts: 1) n nodes send fragmented data to their two neighboring nodes respectively; 2) except for the cluster head node, the remaining n−1 nodes pass the minimum generation the tree path forwards the information to the cluster head node. Therefore, the data communication volume of all nodes in a cluster in the LCSDA protocol $CLCSDA$ is

$$CLCSDA = 2n + n - 1 = 3n - 1 \quad (3)$$

The data communication volume of the three algorithms is shown in Figure 4.6. The figure shows that LCSDA's data communication volume of the protocol is always smaller than that of LCCPDA and CPDA. When n is larger, the superiority of the LCSDA protocol is more prominent.

## 10. Results and Discussion

**Table 4.5: Comparison of performance analysis of proposed model with existing models**

| Protocol | LCSDA | LCCPDA | CPDA |
|---|---|---|---|
| Accuracy | 78% | 69% | 60% |
| Precision | 81% | 72% | 65% |

| | | | |
|---|---|---|---|
| **Recall** | 85% | 77% | 70% |
| **F1-score** | 83% | 74% | 67% |

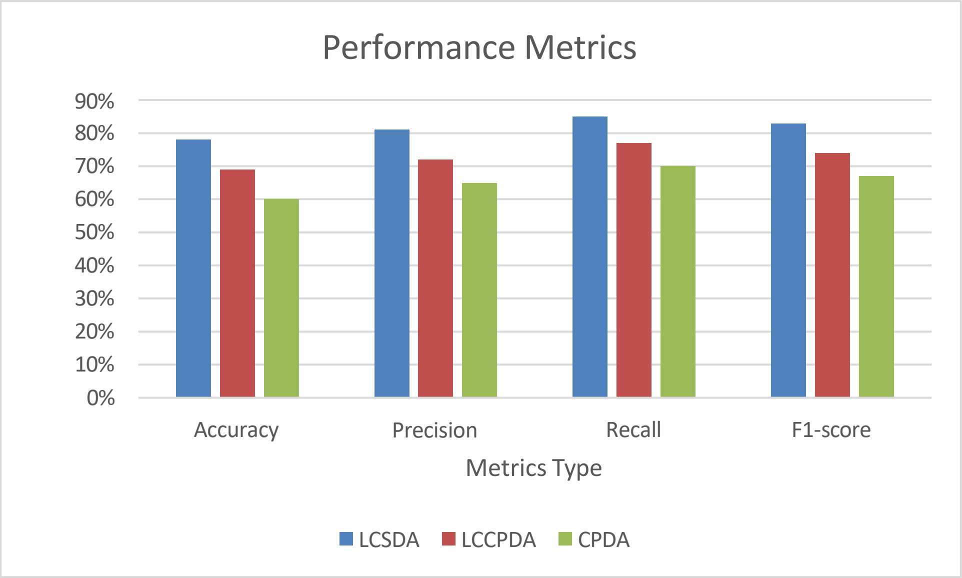

**Figure 4.7:** Performance metrics of proposed model with existing models.

## 11. Conclusion

The research presented in the document aims to develop a low-energy secure data aggregation method (LCSDA) for the Internet of Things (IoT), which balances data privacy, computational efficiency, and energy consumption. It proposes a low-energy-consumption privacy data security fusion method, LCSDA. This method selects two neighbor nodes for each node in the cluster according to the shortest path principle. It uses the Prim minimum spanning tree algorithm to establish the fusion path of the data in the cluster. Experiments show that compared with CPDA and LCCPDA, the LCSDA method can effectively reduce the amount of data communication between nodes while reducing the energy consumption of nodes and the probability of capturing cluster head nodes.

## References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Foster, D. (2019). *Generative deep learning: Teaching machines to paint, write, compose, and play*. O'Reilly Media.

Frustaci, M., Pace, P., Aloi, G., & Fortino, G. (2018). Evaluating critical security issues of the IoT world: Present and future challenges. *IEEE Internet of Things Journal, 5*(4), 2483–2495. https://doi.org/10.1109/JIOT.2017.2767291

García-Magariño, I., Muttukrishnan, R., & Lloret, J. (2019). Human-centric AI for trustworthy IoT systems with explainable multilayer perceptrons. *IEEE Access, 7*, 125562–125574. https://doi.org/10.1109/ACCESS.2019.2937521

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Haria, S. (2019). The growth of the Hide and Seek botnet. *Network Security, 2019*(3), 14–17. https://doi.org/10.1016/S1353-4858(19)30030-5[1]

Hinton, G. E., Krizhevsky, A., & Sutskever, I. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. https://doi.org/10.48550/arXiv.1207.0580

Hodo, E., Edinburgh, X., Branch, J. W., & Getachew, A. T. (2016). Threat analysis of IoT networks using artificial neural network intrusion detection system. In *2016 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISNCC.2016.7750961

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (pp. 448–456). PMLR.

International Data Corporation (IDC). (2019). *The growth in connected IoT devices is expected to generate 79.4ZB of data in 2025, according to a new IDC forecast*.

Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)* (pp. 21–26). ICST.

Kang, K., Pang, Z., Wang, C., & Zhou, L. (2013). Security and privacy mechanism for health internet of things. *The Journal of China Universities of Posts and Telecommunications, 20*(Suppl 2), 64–68. https://doi.org/10.1016/S1005-8885(13)60219-8[1]

Kolosnjaji, B., Eraisha, G., Webster, G., Zarras, A., & Eckert, C. (2016). Deep learning for classification of malware system call sequences. In *Australasian Joint Conference on Artificial Intelligence* (pp. 137–149). Springer. https://doi.org/10.1007/978-3-319-50127-7_12

Kumar, J. S., & Patel, D. (2014). A survey on internet of things: Security and privacy issues. *International Journal of Computer Applications, 90*(14), 20–26. https://doi.org/10.5120/15764-4454

Li, Z., Qin, D., Carlson, M., & Lee, K.-H. (2018). VulDeePecker: A deep learning-based system for vulnerability detection. *arXiv preprint arXiv:1801.01681*. https://doi.org/10.48550/arXiv.1801.01681

Mane, S., & Rao, D. (2021). *Explaining network intrusion detection system using explainable AI framework*. arXiv. https://doi.org/10.48550/arXiv.2103.07110

Marino, D. L., Wickramasinghe, C. S., & Manic, M. (2018). An adversarial approach for explainable AI in intrusion detection systems. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society* (pp. 4183–4188). IEEE. https://doi.org/10.1109/IECON.2018.8592143

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 Military Communications and Information Systems Conference (MilCIS)* (pp. 1–6). IEEE. https://doi.org/10.1109/MilCIS.2015.7348942

Moustafa, N., Slay, J., & Turnbull, B. (2019). An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things. *IEEE Internet of Things Journal, 6*(3), 4815–4830. https://doi.org/10.1109/JIOT.2018.2875467

Pascanu, R., Stokes, J. W., Santhanam, S., & Turner, A. (2015). Malware classification with recurrent networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1916–1920). IEEE. https://doi.org/10.1109/ICASSP.2015.7178318

Suchitra, C., & Vandana, C. P. (2016). Internet of things and security issues. *International Journal of Computer Science and Mobile Computing, 5*(1), 133–139.

da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. (2019). Internet of Things: A survey on machine learning-based intrusion detection approaches. *Computer Networks, 151*, 147–157. https://doi.org/10.1016/j.comnet.2019.01.023

*The Independent*. (2017). Hackers now able to take control of cars to cause deliberate accidents, scientists warn.

Wang, M., Wang, Z., & Ye, X. (2020). An explainable machine learning framework for intrusion detection systems. *IEEE Access, 8*, 73127–73141. https://doi.org/10.1109/ACCESS.2020.2988850

Wang, Z. (2018). Deep learning-based intrusion detection with adversaries. *IEEE Access, 6*, 38367–38384. https://doi.org/10.1109/ACCESS.2018.2852765

Yamaguchi, F., Lindner, F., & Rieck, K. (2011). Vulnerability extrapolation: Assisted discovery of vulnerabilities using machine learning. In *Proceedings of the 5th USENIX Conference on Offensive Technologies* (p. 13). USENIX Association.

Zhao, K., & Ge, L. (2013). A survey on the Internet of Things security. In *2013 Ninth International Conference on Computational Intelligence and Security* (pp. 663–667). IEEE. https://doi.org/10.1109/CIS.2013.145

Zoghi, Z., & Serpen, G. (2021). *UNSW-NB15 computer security dataset: Analysis through visualization*. arXiv. https://doi.org/10.48550/arXiv.2101.05067

García, S., Grill, M., Stiborek, J., & Zunino, A. (2015). An empirical comparison between generative and discriminative classifiers on a malware explanation task. *Virus Bulletin*, 1–8.