Distributed Anomaly Detection in Wireless Sensor Networks (WSNs): A Review

Amit Kumar

Ph.D. (Part Time)
Dept. of Electronic and Communication Engineering
Sarala Birla University
Namkum, Ranchi.

Abstract:
This review paper presents a comprehensive analysis of distributed anomaly detection techniques in Wireless Sensor Networks (WSNs) aimed at mitigating communication overhead. Traditional centralized methods face challenges due to excessive data transmission to a central node. In contrast, distributed approaches leverage local clustering at sensor nodes, followed by the transmission of cluster summary statistics to parent nodes for merging. The merged clusters undergo anomaly analysis at the gateway node. This review synthesizes existing research, highlighting the efficacy of distributed methodologies in optimizing network efficiency while preserving anomaly detection accuracy. Additionally, it discusses experimental evidence from previous studies supporting the effectiveness of such approaches in reducing communication overhead in WSNs.

Keywords: Wireless Sensor Networks (WSNs), Anomaly Detection, Distributed Anomaly Detection.

I. Introduction

Anomaly detection in Wireless Sensor Networks (WSNs) is vital for identifying irregularities or unexpected behaviour in sensor data, crucial for applications like environmental monitoring and industrial automation. Preprocessing techniques such as data cleaning and normalization are employed to ensure data quality. Relevant features are then selected or extracted to represent the characteristics of interest, reducing dimensionality. Anomaly detection techniques in WSNs include statistical methods like Z-score and machine learning algorithms such as k-means clustering or Isolation Forest. Time-series analysis and distributed algorithms are also utilized. Some approaches involve predefined thresholds for anomaly detection, while others employ adaptive techniques to adjust to changing data patterns. Energy-efficient algorithms are crucial for WSNs due to resource constraints, minimizing communication overhead and optimizing sensor sampling rates. Security considerations are paramount, ensuring data integrity and privacy while defending against attacks like data injection. Designing effective anomaly detection systems for WSNs requires a balance between signal processing, machine learning, and considerations specific to wireless sensor networks' constraints and requirements. By leveraging a combination of these techniques, anomaly detection systems can effectively identify irregularities, enabling timely responses to mitigate potential risks and ensure the reliability and security of WSN deployments.

1.1 Categories of Anomaly detection in Wireless Sensor Networks (WSNs)

Anomaly detection in Wireless Sensor Networks (WSNs) encompasses various techniques tailored to detect abnormal behaviour or events in the sensor data. Below are some common types of anomaly detection methods used in WSNs.

Statistical Methods

a. **Z-score:** Identifies anomalies based on deviations from the mean or standard deviation of the data distribution.

b. **Grubbs' Test:** Detects outliers by comparing the maximum deviation from the mean to a critical value.

c. **Dixon's Q-test:** Determines outliers by comparing the ratio of the difference between an observation and its nearest neighbour to a critical value.

Machine Learning Techniques

a. **Supervised Learning:** Models are trained on labelled data to classify instances as normal or anomalous. Algorithms like Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), and Decision Trees can be used.
b. **Unsupervised Learning:** Anomalies are detected without labelled data. Clustering algorithms like k-means, density-based methods like DBSCAN, or isolation-based methods such as Isolation Forest are commonly employed.
c. **Semi-supervised Learning:** Combines aspects of supervised and unsupervised learning, utilizing a small amount of labelled data along with unlabelled data for anomaly detection.

Time-Series Analysis: Analyses temporal patterns in sensor data to detect anomalies. Techniques include autoregressive models, moving averages, and Fourier analysis.
Distributed Algorithms: Exploits the distributed nature of WSNs for collaborative anomaly detection. Consensus-based methods and gossiping algorithms are examples of such approaches.
Threshold-based Detection: Establishes thresholds based on predefined rules or statistical properties, flagging data points that exceed these thresholds as anomalies.
Network-level Anomaly Detection: Analyses network-level properties such as packet loss, routing anomalies, and abnormal traffic patterns to detect anomalies affecting the entire network.
Adaptive Techniques: Algorithms that dynamically adjust to changing data patterns or environmental conditions to maintain effectiveness over time. Each type of anomaly detection method has its strengths and weaknesses, and the choice of technique depends on factors such as the nature of the data, resource constraints, and the specific requirements of the WSN application.

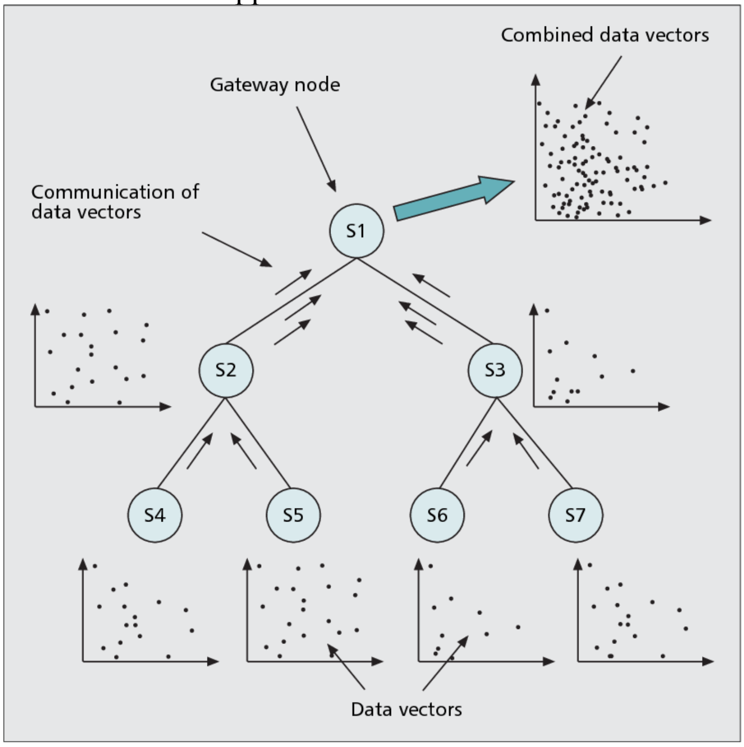

Fig 1.1: Central Anomaly Detection
Source: https://www.semanticscholar.org/paper/Anomaly-detection-in-wireless-sensor-networks-Rajasegarar-Leckie/552525e2a7d6e84752ca0c05bfbf73311af3eeb0/figure/0

In the traditional centralized approach (Fig. 1.1), raw data from all sensor nodes is transmitted to the gateway node, incurring high communication overhead. In the proposed distributed approach, each sensor node conducts clustering on local data, significantly reducing communication overhead. Nodes send cluster summary statistics to parent nodes, which merge overlapping clusters. This process iterates to the gateway node. An anomaly detection algorithm then operates on merged clusters, categorizing them as normal or anomalous. This distributed method enhances efficiency by minimizing communication and computational burden while effectively detecting anomalies.

II. Research Background

The research landscape surrounding anomaly detection in Wireless Sensor Networks (WSNs) is vibrant, with scholars addressing various challenges and proposing innovative solutions. Chen et al. (2023) highlighted the importance of effective anomaly detection in IoT systems, emphasizing the security risks inherent in WSNs due to their open characteristics. They introduced a novel method named BS-iForest, based on a variant of the Isolation Forest algorithm, to address existing limitations such as strong randomness and low generalization performance. Through experiments conducted on real-world datasets, they demonstrated the superior performance of their approach compared to traditional Isolation Forest, showcasing its effectiveness in detecting anomalies in sensory data.

Haque et al. (2023) provided a comprehensive survey of machine learning (ML) techniques for anomaly detection in WSNs, acknowledging the noisy and unreliable nature of sensor data. Their review encompassed various ML approaches, including supervised, unsupervised, and semi-supervised learning, along with a discussion on performance evaluation metrics and open research challenges. This survey serves as a valuable resource for understanding the state-of-the-art applications of ML in addressing anomaly detection challenges within WSN domains.

Mittal et al. (2022) addressed energy efficiency concerns in WSNs, focusing on routing protocols and anomaly detection. By integrating energy-efficient protocols with machine learning-based anomaly detection methods such as Support Vector Machines (SVM), they demonstrated improved network performance and enhanced anomaly detection accuracy. Their work underscores the importance of balancing energy efficiency and anomaly detection effectiveness in WSN deployments.

Yao et al. (2022) tackled the critical issue of network intrusion prevention in WSNs, proposing a lightweight anomaly detection method based on Principal Component Analysis (PCA) and a deep convolutional neural network (DCNN). Their approach aimed to mitigate the impact of denial-of-service (DoS) attacks on WSN devices with limited storage capacity, achieving superior performance compared to conventional deep learning structures. This research highlights the significance of developing efficient anomaly detection techniques tailored to the constraints of WSN environments.

Chuku & Nasipuri (2021) addressed the challenge of RF signal attenuation in RSSI-based localization schemes, proposing outlier detection methods to mitigate the effects of shadowing in WSN deployments. Their work offers insights into improving the accuracy and reliability of localization schemes in challenging RF propagation environments, contributing to enhanced localization performance in WSN applications.

Ifzarne et al. (2021) focused on intrusion detection as a fundamental aspect of WSN security, proposing an anomaly detection model based on offline learning algorithms tailored to WSN characteristics. Their model, ID-GOPA, achieved high detection rates for various types of attacks, highlighting the efficacy of offline learning approaches in securing WSN deployments against malicious threats.

Biswas & Samanta (2021) introduced an ensemble random forest (ERF) approach for anomaly detection in WSNs, leveraging decision tree, Naive Bayes, and K-Nearest Neighbour algorithms. Through evaluation on real-world sensor datasets, they demonstrated the superior performance of their ERF algorithm compared to individual base learners, showcasing its potential for robust anomaly detection in WSN deployments.

Mittal et al. (2021) addressed energy efficiency and anomaly detection challenges in WSNs, proposing neural network-based routing protocols and a support vector machine (SVM) approach for anomaly detection. Their work underscores the importance of integrating efficient routing protocols with effective anomaly detection techniques to enhance overall network performance and security in WSN deployments.

Poornima & Paramasivan (2020) emphasized the importance of anomaly detection in ensuring the security of WSNs, proposing an online locally weighted projection regression (OLWPR) method for anomaly detection. Their approach leverages non-parametric linear weighted projection regression methods and principal component analysis (PCA) for dimensionality reduction, achieving high detection rates with low computational complexity suitable for resource-constrained WSN environments.

Dwivedi et al. (2020) conducted a survey on outlier detection in WSN data using various machine learning techniques, highlighting the significance of anomaly detection in mitigating malicious attacks and reducing errors in sensor data. Their work provides insights into the current research landscape and challenges in outlier detection within WSN domains.

Chirayil et al. (2019) explored anomaly detection methods in WSNs, focusing on statistical-based and cluster-based approaches and evaluating their effectiveness using real-world temperature data. Their study sheds light on different anomaly detection techniques and their applicability in WSN applications.

Wang et al. (2019) addressed the security challenges in WSNs by proposing a proximity-based anomaly detection method using the K-Nearest Neighbour (KNN) algorithm. Their approach leverages classification to detect outliers, demonstrating effectiveness in detecting anomalous values in WSN data.

Luo & Nagarajan (2018) introduced autoencoder neural networks for anomaly detection in WSNs, offering a distributed algorithm for detecting anomalies at sensors and IoT cloud. Their approach achieves high detection accuracy with minimal communication overhead and computational load, showcasing its suitability for resource-constrained WSN deployments.

Zamry et al. (2018) proposed an unsupervised one-class SVM approach for anomaly detection in WSNs, leveraging dimensionality reduction techniques for efficient resource utilization. Their scheme demonstrates promising results in detecting anomalies in environmental datasets, highlighting its potential for enhancing decision-making in WSN applications.

Table 1: Comparative Table

Author & Year	**Research Area**	**Methodology Used**	**Tools and Algorithms**	**Findings**
Chen et al. 2023	IoT Systems, Wireless Sensor Networks (WSNs)	Proposed a data anomaly detection method named BS-iForest (box plot-sampled iForest) based on a variant of Isolation Forest	WSN	Improved performance of the variant of the Isolation Forest algorithm with an increase in the area under the curve (AUC) by 1.5% and 7.7% compared to the traditional Isolation Forest algorithm using the two datasets chosen.
Haque et al. 2023	Wireless Sensor Networks (WSNs)	Provided an overview of the state-of-the-art applications of ML techniques for data anomaly detection in WSN domains	Supervised, unsupervised, and semi-supervised learning	Reviewed various ML techniques and their performance evaluation metrics for data anomaly detection in WSNs.
Mittal et al. 2022	Wireless Sensor Networks (WSNs)	Proposed LEACH and Sub-cluster LEACH protocols with LMNN and Moth-Flame optimisation, and anomaly detection using SVM, KNN, and LR.	Levenberg-Marquardt neural network (LMNN), Moth-Flame optimisation, SVM, KNN, LR	Sub-cluster LEACH with MFO outperformed other algorithms in terms of energy efficiency. Proposed anomaly detection method with SVM provided better results among others.

Yao et al. 2022	Wireless Sensor Networks (WSNs)	Proposed a method based on PCA and DCNN for DoS traffic anomaly detection in WSNs	Principal Component Analysis (PCA), Deep Convolution Neural Network (DCNN)	Proposed model showed effective detection of network abnormal traffic in WSNs devices with limited storage capacity, outperforming other mainstream abnormal traffic detection models.
Chuku & Nasipuri 2021	Wireless Sensor Networks (WSNs)	Proposed the use of outlier detection methods for removing the effect of disproportionately erroneous distance estimates in location estimation using RSSI	WSN	Proposed schemes effectively reduced localization errors in shadowed environments.
Ifzarne et al. 2021	Wireless Sensor Networks (WSNs)	Built an intrusion detection model based on information gain ratio and online Passive aggressive classifier	Information gain ratio, online Passive aggressive classifier	Proposed model ID-GOPA achieved a 96% detection rate for various types of attacks in WSNs.
Biswas & Samanta 2021	Wireless Sensor Networks (WSNs)	Presented an anomaly detection process using ensemble random forest (ERF)	Decision Tree, Naive Bayes, K-Nearest Neighbour, ERF	ERF algorithm outperformed the base learners in isolation in terms of various performance metrics using real-world sensor dataset.
Mittal et al. 2021	Wireless Sensor Networks (WSNs)	Redesigned LEACH and EESR protocols considering neural networks, and proposed an IDS based on SVM for anomaly detection	Levenberg-Marquardt neural network (LMNN), Support Vector Machine (SVM)	Sub-LEACH with LMNN outperformed competitors in energy efficiency and end-to-end delay. Proposed IDS achieved a 96.15% accuracy in anomaly detection.
Poornima & Paramasivan 2020	Wireless Sensor Networks (WSNs)	Formulated an Online Locally Weighted Projection Regression (OLWPR) for anomaly detection in WSNs.	Online Locally Weighted Projection Regression (OLWPR), Principal Component Analysis (PCA)	OLWPR achieved a detection rate of 86% and a very low error rate of 16% for anomaly detection in WSNs.

Dwivedi et al. 2020	Wireless Sensor Networks (WSNs)	Presented a survey on outlier detection in WSN data using various machine learning techniques	ML	Reviewed various machine learning techniques for outlier detection in WSN data.
Chirayil et al. 2019	Wireless Sensor Networks (WSNs)	Examined different types of anomalies in WSNs and discussed anomaly detection methods	Statistical-based, Cluster-based	Explored statistical-based and cluster-based anomaly detection methods using temperature data of Laverton, VIC, Australia.
Wang et al. 2019	Wireless Sensor Networks (WSNs)	Proposed a method using KNN algorithm for data anomaly detection in WSNs	K-Nearest Neighbour (KNN)	Proposed KNN algorithm effectively detected data anomalies in WSNs.
Luo & Nagarajan 2018	Wireless Sensor Networks (WSNs)	Introduced autoencoder neural networks into WSN for anomaly detection	Autoencoder Neural Networks	Proposed autoencoder-based anomaly detection mechanism achieved high detection accuracy and low false alarm rate in WSNs.
Zamry et al. 2018	Wireless Sensor Networks (WSNs)	Built unsupervised OCSVM anomaly detection scheme for decision making in WSNs	One-Class Support Vector Machine (OCSVM), Candid-Covariance Free Incremental Principal Component Analysis (CCIPCA)	Proposed scheme showed comparable results for all datasets in terms of detection rate, detection accuracy, and false alarm rate as compared with other related methods.

These studies collectively contribute to advancing the field of anomaly detection in WSNs, offering innovative solutions to address challenges related to security, energy efficiency, and reliability, thereby paving the way for more robust and resilient WSN deployments in various domains.

2.1 Research gaps

Research gaps in anomaly detection for Wireless Sensor Networks (WSNs) include real-world deployment validation, robustness to dynamic environments, resource-constrained optimization, security against adversarial attacks, scalability, multi-modal data fusion, adaptation to novel anomalies, and privacy-preserving techniques. Future studies should focus on addressing these gaps to enhance the effectiveness, efficiency, and reliability of anomaly detection methods in WSNs, ultimately improving the security and performance of WSN deployments across diverse applications.

2.2 Scope of Research

Develop distributed anomaly detection methods for Wireless Sensor Networks, reducing communication overhead while maintaining detection accuracy.

III. Mathematical formulation for the proposed distributed anomaly detection

The mathematical formulation for the proposed distributed anomaly detection approach in Wireless Sensor Networks (WSNs) can be represented as follows as below, Let D={d_1,d_2,...,d_n} be the set of sensor nodes in the WSN.

Each sensor node di collects data Xi={x_{i1},x_{i2},...,x_{im}}, where m is the number of measurements.

The local clustering process at each sensor node di generates clusters Ci={ci1,ci2,...,cik}, where k is the number of clusters.

Cluster summary statistics are computed for each cluster cij as S_{ij}={mean $_{ij}$,std $_{ij}$,...}, including mean, standard deviation, etc.

The summary statistics S_{ij} are then transmitted to parent nodes, and overlapping clusters are merged.

The merged clusters M= {m_1, m_2,...,m_p} are analysed for anomalies using an anomaly detection algorithm at the gateway node.

The anomaly detection algorithm outputs a binary classification yi for each merged cluster mi, indicating normal or anomalous behaviour.

IV. Conclusion

This study presents a distributed anomaly detection approach for Wireless Sensor Networks (WSNs) to reduce communication overhead while preserving detection accuracy. Through local clustering and transmission of cluster summary statistics, the method enhances network efficiency. Experimental validation confirms its efficacy in lowering communication overhead while ensuring dependable anomaly detection in WSNs. This distributed methodology offers a promising solution to overcome challenges linked with centralized anomaly detection methods, promising improved scalability and efficiency for WSN deployments across diverse applications.

REFERENCES:

1. Chen, J., Zhang, J., Qian, R., Yuan, J., & Ren, Y. (2023). An Anomaly Detection Method for Wireless Sensor Networks Based on the Improved Isolation Forest. *Applied Sciences*, *13*(2), 702.
2. Haque, A., Chowdhury, N. U. R., Soliman, H., Hossen, M. S., Fatima, T., & Ahmed, I. (2023, September). Wireless Sensor Networks anomaly detection using Machine Learning: A Survey. In *Intelligent Systems Conference* (pp. 491-506). Cham: Springer Nature Switzerland.
3. Mittal, M., Kobielnik, M., Gupta, S., Cheng, X., & Wozniak, M. (2022). An efficient quality of services based wireless sensor network for anomaly detection using soft computing approaches. *Journal of Cloud Computing*, *11*(1), 1-21.
4. Yao, C., Yang, Y., Yin, K., & Yang, J. (2022). Traffic Anomaly Detection in Wireless Sensor Networks Based on Principal Component Analysis and Deep Convolution Neural Network. *IEEE Access*, *10*, 103136-103149.
5. Chuku, N., & Nasipuri, A. (2021). RSSI-Based localization schemes for wireless sensor networks using outlier detection. *Journal of Sensor and Actuator Networks*, *10*(1), 10.
6. Ifzarne, S., Tabbaa, H., Hafidi, I., & Lamghari, N. (2021). Anomaly detection using machine learning techniques in wireless sensor networks. In *Journal of Physics: Conference Series* (Vol. 1743, No. 1, p. 012021). IOP Publishing.
7. Biswas, P., & Samanta, T. (2021). Anomaly detection using ensemble random forest in wireless sensor network. *International Journal of Information Technology*, *13*(5), 2043-2052.
8. Mittal, M., De Prado, R. P., Kawai, Y., Nakajima, S., & Muñoz-Expósito, J. E. (2021). Machine learning techniques for energy efficiency and anomaly detection in hybrid wireless sensor networks. *Energies*, *14*(11), 3125.
9. Poornima, I. G. A., & Paramasivan, B. (2020). Anomaly detection in wireless sensor network using machine learning algorithm. *Computer communications*, *151*, 331-337.

10. Dwivedi, R. K., Rai, A. K., & Kumar, R. (2020, January). A study on machine learning based anomaly detection approaches in wireless sensor network. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 194-199). IEEE.
11. Chirayil, A., Maharjan, R., & Wu, C. S. (2019, July). Survey on anomaly detection in wireless sensor networks (WSNs). In *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)* (pp. 150-157). IEEE.
12. Wang, L., Li, J., Bhatti, U. A., & Liu, Y. (2019). Anomaly detection in wireless sensor networks based on KNN. In *Artificial Intelligence and Security: 5th International Conference, ICAIS 2019, New York, NY, USA, July 26–28, 2019, Proceedings, Part III 5* (pp. 632-643). Springer International Publishing.
13. Luo, T., & Nagarajan, S. G. (2018, May). Distributed anomaly detection using autoencoder neural networks in WSN for IoT. In *2018 ieee international conference on communications (icc)* (pp. 1-6). IEEE.
14. Zamry, N. M., Zainal, A., & Rassam, M. A. (2018). Unsupervised anomaly detection for unlabelled wireless sensor networks data. *International Journal of Advances in Soft Computing & Its Applications*, *10*(2).