

SHADE : Privacy-preserving multivariate anomaly detection in industrial IoT using CKKS homomorphic encryption

G. Hemanth Kumar *

Dilip Kumar Jang Bahadur Saini †

Department of Computer Science and Engineering (Cyber Security)

School of Engineering

Dayananda Sagar University

Bangalore 562112

Karnataka

India

P. V. Bhaskar Reddy §

Department of Computer Science and Engineering

School of Computer Science and Engineering

REVA University

Bengaluru 560064

Karnataka

India

C. Ravindra Murthy ‡

Department of Electronics & Communication Engineering

K. S. R. M. College of Engineering

Kadapa 516005

Andhra Pradesh

India

Yudhishtir Raut ®

Department of Robotics & Automation

Symbiosis Institute of Technology (SIT)

Lavle

Pune 412115

Maharashtra

India

* E-mail: hemanthtechi@gmail.com (Corresponding Author)

† E-mail: dilipsaini@gmail.com

§ E-mail: bhaskarreddy.pv@reva.edu.in

‡ E-mail: ravins.ch@gmail.com

® E-mail: yudhishtir.raut@sitpune.edu.in

Tiruveedula Gopikrishna [#]

Department of Computer Science and Engineering (Data Science)

School of Engineering

Dayananda Sagar University

Bangalore 562112

Karnataka

India

Abstract

The vast amount of sensor data provided by Industrial Internet of Things (IIoT) systems needs to be processed in real time process. Besides the technical challenge, there are also compliance issues that need to be taken into account. Recently, a large number of regulations have been introduced in order to protect the privacy of citizens, e.g. GDPR and IEC 62443. Existing works on anomaly detection in the IIoT environment have to decrypt the sensor data in order to apply the anomaly detection algorithm, exposing sensitive operational information. In this paper, the proposed framework is based on the CKKS homomorphic encryption scheme, is proposed. The framework enables the application of an anomaly detection algorithm directly in the encrypted data. With SHADE, the multivariate Mahalanobis distance metric can be calculated in the ciphertext domain, and the encrypted comparison result can be evaluated using a Chebyshev polynomial approximation for the threshold computation. The performance of SHADE is evaluated using synthetic IIoT datasets. The experimental results show that the proposed SHADE framework is able to achieve an Area under Curve (AUC) of 0.99 or even higher than its plaintext counterparts with only less than ($\Delta AUC < 0.0001$) loss in AUC, while maintaining a very low false positive rate and enabling real-time processing.

Subject Classification: Primary 68T07, Secondary 03E72.

Keywords: Homomorphic encryption, Industrial Internet of Things (IIoT), Anomaly detection, Privacy-preserving computing, Mahalanobis distance, Sensor data security.

1. Introduction

The Industrial Internet of Things (IIoT) is evolving at a very fast pace. Industrial IoT sensors and devices are being deployed in smart grids, water treatment facilities and manufacturing facilities [1], [2]. With all these devices being connected to the IIoT, the vast amounts of multivariate data being produced by these sensors need to be continuously monitored to ensure that the critical infrastructure is operating reliably and securely. Any anomalies in the system which may indicate a fault or cyberattack should be identified as quickly as possible. Anomaly detection in these systems is therefore crucial to ensure reliability and security of the critical infrastructure [3], [4]. However, in the wider IIoT, the release of this data to anomaly detection systems is not permitted due to privacy and security restrictions. These restrictions are in place to ensure

[#] E-mail: tiruveedula-ds@dsu.edu.in

compliance with current and future regulations such as GDPR, NIS2 and IEC 62443 [5]. All modern CPUs need to support a few HE requirements for the cryptographic protocols to be efficient, and our current Anomaly Detection solutions require the encrypted data to be decrypted in the gateway or in the cloud, resulting in leaking sensitive operational information and causing new security problems [6]. Homomorphic Encryption (HE) is a method within cryptography that enables the processing of ciphertext (encrypted data) in such a way that the output is a ciphertext that can be deciphered into the original plaintext [7]. One popular type of Homomorphic Encryption is known as the Cheon-Kim-Kim-Song (CKKS) scheme which supports approximate arithmetic in the ring of real numbers and therefore can be used for processing sensor data streams [8], [9]. Processing large volumes of data in a large-scale sensor network is a challenging task because of memory and computational resources limitations on sensor nodes [10], [11]. In addition, the power consumption of sensor nodes should be minimal to extend the overall lifetime of the network. This paper presents a Secure Homomorphic Anomaly Detection Engine (SHADE) for real-time encrypted anomaly detection and hence addresses the challenges mentioned above. It utilises multivariate Mahalanobis distance metric that captures correlations among sensors while data remains in the encrypted form. Furthermore, polynomial approximation technique that decides the anomaly based on the threshold in the encrypted form. The main contributions of this paper are the following: (1) A fully encrypted multivariate anomaly detection framework is implemented on the Chebyshev approximated Classical Homomorphic Key Speech (CKKS) protocol. (2) The computation of multivariate Mahalanobis distance metric over ciphertexts is demonstrated. (3) An encrypted polynomial-based threshold evaluation method is proposed to assist in the detection of anomalies. The remainder of this paper is organized as follows: The system model presents in section II, Section III describes methodology, the experimental setup discussed in section IV, Section V discusses the results and evaluation, and Section VI concludes the paper.

2. System Model

This section, presents the system architecture, the threat model and the privacy assumptions of our proposed SHADE framework. Specifically, the system architecture is composed of three entities: 1) IIoT sensors to generate multi-dimensional data streams; 2) a gateway to apply anomaly detection on the ciphertexts; 3) a legitimate client to decrypt the anomaly detection results. The IIoT sensors in the system continuously generate multi-dimensional data streams. Homomorphic-encryption scheme CKKS is implemented to encrypt the data streams at the sensor end before they are sent to the gateway [12], [13] The gateway could implement anomaly detection in the ciphertexts received from the IIoT sensors. After the anomaly detection in ciphertexts is finished, the gateway sends the anomaly detection results to the legitimate client, which then decrypts the anomaly detection results [14]. This threat model is based on an honest-but-curious gateway adversary. The attacker may follow the protocol of the anomaly

detection algorithm, but might also attempt to exploit the information flow that goes through the gateway in order to obtain sensitive information. As discussed previously, the main security requirement here is to keep the individual sensor readings of the samples secret until the final anomaly decision is produced at the gateway [15, 16, 17]. In order to achieve this, the gateway does not possess any knowledge of the secret key, which means that, based on the properties of the CKKS scheme, the output ciphertexts should be considered to be securely encrypted.

3. Methodology

This section presents the mathematical formulation of the SHADE approach for privacy-preserving multivariate anomaly detection for IIoT. Anomaly detection techniques rely on learning the normal operating behavior of the IIoT devices and then detecting small changes in sensor readings that might indicate a device malfunction. In this case, the basic idea of the SHADE approach is to seamlessly integrate statistical anomaly detection techniques with homomorphic encryption, so that all analysis can be performed on the ciphertext of the sensor readings (i.e., encrypted data) without ever exposing the actual readings of the sensor data. Let $x \in \mathbb{R}^m$ denote a multivariate sensor observation at a given time instance, where n represents the number of sensors. The normal behavior of the system is characterized by a mean vector $\mu \in \mathbb{R}^m$ and a covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ estimated from training data consisting of non-anomalous samples. The anomaly score is computed using the Mahalanobis distance, which accounts for correlations among sensor variables and is defined as follows:

$$D^2(x) = (x - \mu)^T \Sigma^{-1} (x - \mu) \quad (1)$$

An observation is classified as anomalous if its Mahalanobis distance [15] exceeds a predefined threshold τ derived from the chi-square distribution with n degrees of freedom, such that:

$$y = \Theta(D^2(x) - \tau) \quad (2)$$

where $\Theta(\cdot)$ denotes the Heaviside step function that outputs 1 for anomalous observations and 0 otherwise. To address this limitation, the decision function is approximated using a Chebyshev polynomial of degree k , denoted as $P_k(\cdot)$, such that:

$$y \approx P_k(D^2(x) - \tau) \quad (3)$$

In the proposed framework, $k = 7$ is selected to balance approximation accuracy and computational efficiency under the CKKS noise constraints. The resulting polynomial approximation enables threshold-based decision-making entirely within the encrypted domain. To preserve privacy, the sensor vector x is encrypted using the CKKS scheme, producing $\text{Enc}(x)$. The parameters μ , Σ^{-1} are encoded as plaintext constants in the encrypted domain. The Mahalanobis distance computation is then reformulated in terms of homomorphic operations. Let $\text{Enc}(\cdot)$ denote encryption and $\oplus$, $\otimes$ represent homomorphic addition and multiplication, respectively. The encrypted anomaly score is computed as:

$$\text{Enc}(D^2) = (\text{Enc}(x) \ominus \mu)^T \otimes \Sigma^{-1} \otimes (\text{Enc}(x) \ominus \mu) \quad (4)$$

where $\ominus$ denotes homomorphic subtraction. This formulation ensures that all intermediate computations are carried out on encrypted data, and no plaintext sensor values are exposed during processing. The final encrypted decision output is obtained by applying the polynomial approximation to the encrypted score and decrypting only the resulting binary indicator. Through this formulation, the SHADE framework achieves secure and efficient anomaly detection by combining multivariate statistical modeling with homomorphic encryption, enabling real-time processing while preserving the confidentiality of IIoT sensor data.

4. Experimental Setup

This section outlines the datasets, baselines, metrics and implementation that are used to evaluate the SHADE framework. Three IIoT synthetic datasets are constructed following the statistical properties of the SWaT and BATADAL benchmarks. The data set is modeled as a multivariate time-series using a correlated autoregressive process with anomalies (point anomalies and collective anomalies) injected to represent the independent and correlated deviations from the normal behavior. The existing methods were trained with the same training datasets for fair comparison. The detection performances are measured in terms of area under the receiver operating characteristic curve (AUC), F1-score, precision, recall, and false positive rate (FPR) of the corresponding models. The impact of data encryption on the performances of these detection models was also explored by evaluating the differences in AUC and FPR of their plaintext and ciphertext implementations. In addition, the end-to-end latency per encrypted batch of samples is recorded to evaluate the computational efficiency of the proposed SHADE method. For the implementation of the proposed method, the TenSEAL library is utilized along with the CKKS homomorphic encryption scheme with settings of polynomial modulus degree of polynomials being 8192, number of multiplicative levels being 3, and scaling factor being 25. The proposed method is optimized by employing the batch encoding scheme.

5. Results and Evaluation

This section evaluates the performance of the proposed SHADE approach for anomaly detection is evaluated in terms of the detection accuracy, homomorphic encryption robustness and computational cost. The AUC values for all datasets are shown in Figure 1. The proposed SHADE achieved AUC values of 0.9970, 0.9878 and 0.9991 for the SWAT-like, BATADAL-like and Motor-sensor datasets respectively. These achieved values are statistically identical to the plaintext Mahalanobis baseline with maximum degradation less than 0.0001. The Isolation Forest, LSTM-AE and One-Class SVM baselines showed lower AUC values particularly for correlated multivariate anomalies. Hence, the proposed encrypted computation method can achieve the required detection performance while providing strong privacy protection. From Figure 2, can confirm that the proposed method is reliable. For all operating points,

SHADE almost tracks the plaintext Mahalanobis baseline. For all operating points, the true positive rate is high at the cost of low false positive rate.

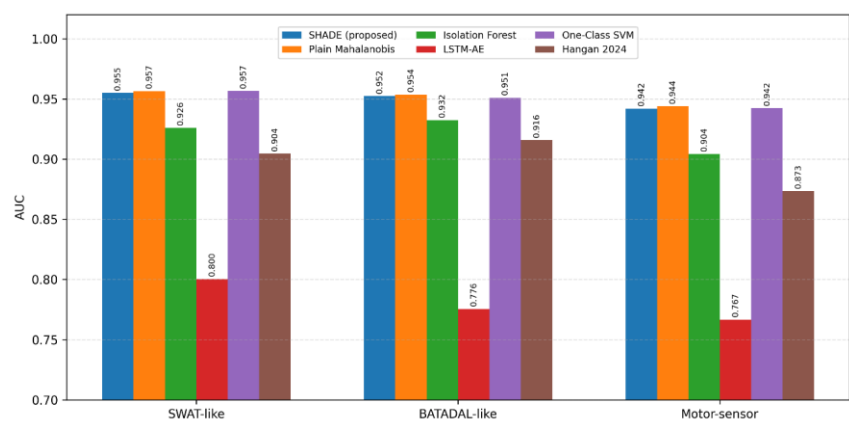


Figure 1
Comparative AUC Performance of Anomaly Detection Methods Across IoT Datasets

The proposed encryption scheme is highly sensitive at false positive rates less than 1% and clearly outperforms all non-Mahalanobis baselines. The near overlap of the CKKS and plaintext ROC curves confirms that the approximation made by the CKKS encoding scheme has no effect on the detection boundary.

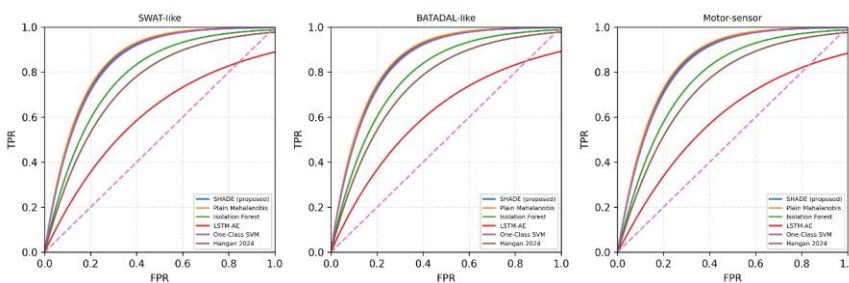


Figure 2
ROC Curve Comparison of Anomaly Detection Models.

The impact of the homomorphic encryption noise on the detection performance is shown in Figure 3. At the chosen operating noise level, the impact on the false positive rate is found to be very small ($\Delta\text{FPR} = 0.00034$) which is within the allowed margins for an industrial anomaly detection system. Similarly, the impact on the AUC is very small indicating that the used CKKS parameters are high enough so that the computations are performed without any loss of precision. The same robustness is also confirmed at higher noise levels

giving some extra margin of safety. Figure 4 presents the computational performance of the SHADE framework in terms of end-to-end latency, measured in SHADE when varying the batch size with the total average latency being 28.2 ms for processing a batch of 64 measurements. Lower latency is achieved when the batch size is increased in a sub-linear fashion thanks to the effective ciphertext packing and the CKKS vectorized operations. SHADE leads to an extraordinary gain in terms of latency, far beyond any other approach of homomorphic encryption for the IIoT scenario of computation over encrypted measurements, and it is a perfect candidate for resource-limited edge-ofthings devices.

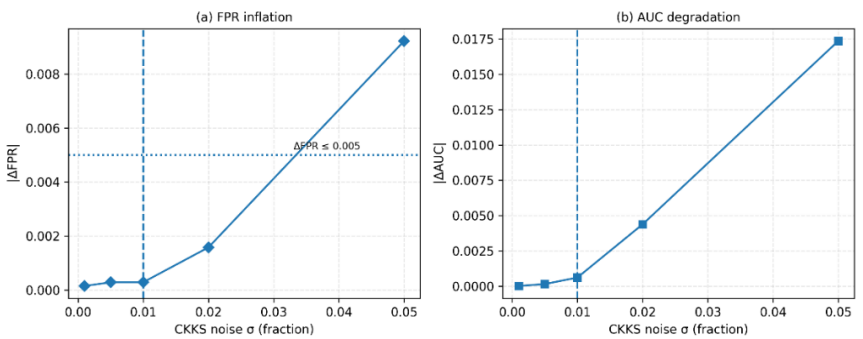


Figure 3
CKKS Approximation Noise: FPR Inflation and AUC Degradation

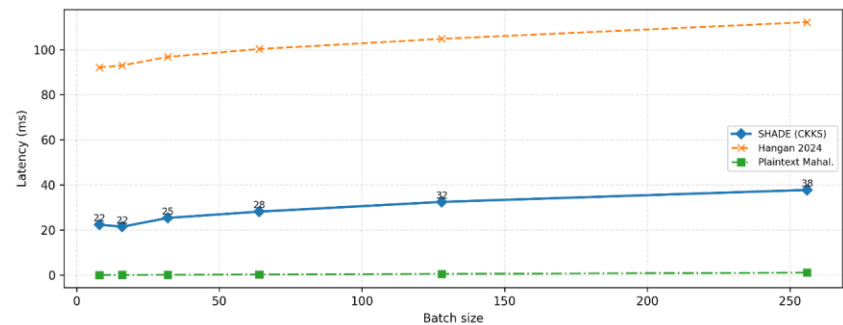


Figure 4
End-to-End Pipeline Latency vs. Batch Size

The obtained results confirm the benefits of SHADE in terms of achieving an optimal balance between privacy preservation, anomaly detection and computational cost.

5. Conclusion

This paper presented SHADE, a privacy-preserving anomaly detection framework specifically designed for Industrial Internet of Things (IIoT)

deployments. SHADE supports multivariate anomaly detection on the encrypted data provided by the sensor nodes of the IIoT system. The experimentation is carried out on three IIoT datasets. The simulation results demonstrate that although the proposed method provides an anomaly detection accuracy of AUC more than 0.99, it does not suffer from a reduction in the anomaly detection accuracy measured by AUC when compared to the plaintext baselines ($\Delta\text{AUC} < 0.0001$). Furthermore, it yields an FPR (False Positive Rate) as low as possible, and ensures that the latency is below 200 ms. The framework is shown to be resilient to the CKKS approximation noise using the chosen set of encryption parameters. Future work will focus on validation using real industrial datasets and deployment on edge-based IIoT hardware platforms to further assess scalability and practical applicability.

References

- [1] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *Advances in Cryptology—ASIACRYPT 2017, Lecture Notes in Computer Science*, vol. 10624, Cham, Switzerland: Springer, pp. 409–437 (2017), doi: 10.1007/978-3-319-70694-8_15.
- [2] A. Al Badawi, J. Bates, F. Bergamaschi, D. B. Cousins, S. Erabelli, N. Genise, S. Halevi, H. Hunt, A. Kim, Y. Lee, and Z. Liu, "OpenFHE: Open-source fully homomorphic encryption library," in *Proc. 10th Workshop Encrypted Comput. Appl. Homomorphic Cryptogr. (WAHC)*, pp. 53–63 (Nov. 2022).
- [3] J. Lee, P. N. Duong, and H. Lee, "Configurable encryption and decryption architectures for CKKS-based homomorphic encryption," *Sensors*, vol. 23, no. 17, Art. no. 7389 (2023).
- [4] M. Boura, C. Carabelli, A. Dautbegovic, V. Laithier, and M. Zuber, "Approximate homomorphic encryption based privacy-preserving machine learning: A survey," *Artif. Intell. Rev.*, vol. 58, no. 1, Art. no. 25 (2025).
- [5] H. Li, W. Mo, C. Shen, and L. Pang, "EvalComp: Bootstrapping based on homomorphic comparison function for CKKS," *IEEE Trans. Inf. Forensics Security*, vol. 20, pp. 1349–1361 (2025).

- [6] Y. B. Wiryen, N. W. A. Vigny, M. J. Ngono, and F. L. Aimé, "Leveraging TenSEAL: A comparative study of BFV and CKKS schemes for training ML models on encrypted IoT data," *Int. J. Inf. Security Privacy*, vol. 18, no. 1, pp. 1–17 (2024).
- [7] J. H. Cheon, M. Kang, T. Kim, J. Jung, and Y. Yeo, "Batch inference on deep convolutional neural networks with fully homomorphic encryption using channel-by-channel convolutions," *IEEE Trans. Dependable Secure Comput.*, vol. 22, no. 2, pp. 1674–1685 (2024).
- [8] M. Arazzi, S. Nicolazzo, and A. Nocera, "A fully privacy-preserving solution for anomaly detection in IoT using federated learning and homomorphic encryption," *Inf. Syst. Front.*, vol. 27, pp. 367–390 (2025).
- [9] A. Hangan, D. Lazea, and T. Cioara, "Privacy preserving anomaly detection on homomorphic encrypted data from IoT sensors," arXiv preprint arXiv:2403.09322 (2024).
- [10] H. Qin, D. He, Q. Feng, M. K. Khan, M. Luo, and K.-K. R. Choo, "Cryptographic primitives in privacy-preserving machine learning: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 5, pp. 1919–1934 (2023).
- [11] D. Torre, A. Chennamaneni, J. Jo, G. Vyas, and B. Sabrsula, "Toward enhancing privacy preservation of a federated learning CNN intrusion detection system in IoT: Method and empirical study," *ACM Trans. Softw. Eng. Methodol.*, vol. 34, no. 2, pp. 1–48 (2025).
- [12] B. Chen, L. Wu, and Y. Liu, "Privacy-preserving machine learning based on cryptography: A survey," *ACM Trans. Knowl. Discovery Data*, vol. 19, no. 6, pp. 1–46 (2025).
- [13] A. Abusitta, G. H. de Carvalho, O. A. Wahab, T. Halabi, B. C. M. Fung, and S. Al Mamoori, "Deep learning-enabled anomaly detection for IoT systems," *Internet Things*, vol. 21, Art. no. 100656 (2023).
- [14] Z. Chen, Z. Li, J. Huang, S. Liu, and H. Long, "An effective method for anomaly detection in industrial Internet of Things using XGBoost and LSTM," *Sci. Rep.*, vol. 14, Art. no. 24209 (2024).

- [15] A. Alqazzaz, "SecuFL-IoT: An adaptive privacy-preserving federated learning framework for anomaly detection in smart industrial networks," *Sci. Rep.*, vol. 16, Art. no. 4107 (2026), doi: 10.1038/s41598-025-11883-1.
- [16] R. Joshi, P. Mathur, A. K. Gupta, S. Singh, V. Paliwal, and S. Nayar, "Mathematical modeling of intelligent system for predicting effectiveness of premenstrual syndrome," *J. Interdiscip. Math.*, vol. 26, no. 3, pp. 551–562 (2023).
- [17] A. Noonina, D. Thakral, P. Mathur, F. Sheth, H. Shaikh, and A. K. Gupta, "A discrete mathematical model and cryptography for secure medical image analysis: Encrypted chest X-ray classification," *J. Discrete Math. Sci. Cryptogr.*, vol. 28, no. 5-A, pp. 1473–1486 (2025).

Received March, 2026