LLM-Based Detection of Cyber Anomalies in Industrial Control Systems

Prabhav V Thali
Department of Computer Science & Engineering
COEP Technological University
Pune, India
thalip23.comp@coeptech.ac.in

Dr. Vinod Pachghare
Department of Computer Science & Engineering
COEP Technological University
Pune, India
vkp.comp@coeptech.ac.in

***Abstract*—The increasing integration of Industrial Control Systems (ICS) with modern networked infrastructures has significantly heightened their susceptibility to advanced cyber threats. Ensuring the resilience and reliability of such critical systems calls for advanced anomaly detection solutions capable of promptly identifying and mitigating malicious behavior. Traditional log-based detection approaches using machine learning typically rely on expert-crafted features and often lack adaptability across varied ICS environments. In this study, we present a language model-based anomaly detection framework that leverages BERT to learn contextual relationships within multivariate ICS log data. The proposed method employs a streamlined preprocessing phase to transform structured system records—comprising sensor outputs, actuator signals, and protocol-level parameters—into sequential inputs optimized for transformer-based models. This architecture is evaluated against established baselines, including a recurrent neural network (RNN) and a TF-IDF plus Support Vector Machine (SVM) model, using a detailed dataset that spans multiple cyberattack scenarios such as network probing, unauthorized command execution, and manipulation of control variables. Our framework demonstrates strong generalization and detection capabilities, achieving an F1-score exceeding 0.80 while maintaining a low false alarm rate. These findings highlight the practicality and effectiveness of applying pre-trained language models to ICS security, offering a scalable and context-aware solution for detecting anomalies in complex operational environments.**

***Index Terms*—Industrial Control Systems (ICS), Anomaly Detection, Large Language Models, BERT, Cyber Physical Security, Machine Learning.**

I. Introduction

Contemporary industrial control systems (ICS) are fundamental to the operation of critical infrastructure—including energy distribution, water treatment plants, and manufacturing facilities [23]. While the integration of networked sensors and controllers improves efficiency, it also exposes ICS to cyber attacks with potentially catastrophic effects [1]. Highprofile incidents like Stuxnet and Ukranian power grid attacks have demonstrated the real-world impact of ICS intrusions [20]. A 2021 report found that 33.8% of ICS computers were exposed to malicious activity, underscoring the urgent need for improved security monitoring [6]. Anomaly detection has emerged as a key strategy for securing ICS environments, enabling the identification of both known and previously unseen threats by learning deviations from the system's learned behavioral norms, rather than relying solely on predefined attack signatures [1] [2]. Conventional detection techniques—such as rule-based approaches and standard machine learning algorithms—often depend heavily on manual feature extraction to spot irregularities. The dependence of traditional approaches on handcrafted features hinders their scalability across varied ICS environments and limits their ability to detect subtle or sophisticated threats.

Latest enhancements in language modeling has introduced latest capabilities for processing sequential data, which demonstrates strong contextual understanding without depending on additional parsing and domain-specific feature engineering. Models such as BERT have been effective in system log analysis by capturing semantic dependencies directly from raw inputs [2] [3]. Despite these advancements, their integration into ICS cybersecurity particularly for analyzing multivariate logs comprising both sensor and control data remains relatively unexplored. This research proposes a novel anomaly detection framework built on a Large Language Model (LLM), specifically BERT, tailored to identify dormant patterns in ICS logs. Structured telemetry data is transformed into tokenized sequences, enabling the model to learn behavioral context and detect between normal and malicious activity. The proposed method is benchmarked against LSTM and classical machine learning baselines such as TF-IDF combined with Support Vector Machines (SVM). Experimental analysis conducted across various ICS attack types such as command injection, scanning, and control manipulation demonstrate that the LLM-based framework surpasses traditional techniques in both detection accuracy and false positive reduction. This highlights its potential as a scalable and adaptable solution for enhancing anomaly detection in modern industrial environments.

II. Related Work

A. ICS Anomaly Detection Techniques

Early studies focussed on deploying classical machine learning methods, such as Support Vector Machines (SVM), decision tree classifiers, and clustering approaches, leveraging features derived from network traffic data or sensor log entries [19]. While these methods achieved reasonable performance in detecting known attack patterns, their reliance on handcrafted features and domain knowledge limited scalability and robustness to new or stealthy attacks.

To address these limitations, researchers have increasingly adopted deep learning techniques. Sequence-based models like Long Short-Term Memory (LSTM) networks have been employed to understand temporal relationships in multivariate sensor data [18]. By learning from past trends, these models forecast future behavior and identify substantial deviations as potential anomalies [18]. Reconstruction-based models, including autoencoders and variational autoencoders, have also been deployed to model normal ICS behavior, identifying anomalies through reconstruction error thresholds [7]. Ortega-Fernandez showcased the use of a deep autoencoder for detecting network-based DDoS attacks in ICS, achieving strong performance with a low false positive rate [7].

More recently, Abshari et al. [1] proposed AnomalyLLM, which uses Retrieval-Augmented Generation (RAG) leveraging Large Language Models to extract physical invariants from system documentation automatically [1]. This approach reduced false positives and improved detection accuracy. Unlike prior works, our study does not rely on external documentation or rule mining but instead leverages a transformer-based model trained directly on structured ICS logs to learn implicit behavioral patterns.

B. Log Anomaly Detection Using NLP and LLMs

In the domain of system log analysis, traditional methods have relied on template extraction, term frequency techniques, or dimensionality reduction methods such as TF-IDF + SVM and PCA [15]. These approaches perform well under structured log formats but often fail in dynamic environments where logs vary significantly or contain unstructured text.

Recent advances in transformer-based models, have demonstrated strong capabilities in modeling sequential and contextual patterns in log data. The LAnoBERT model, introduced by Lee et al. [3], employs an unsupervised framework to detect anomalies by modeling log sequences. Rather than depending on explicit template extraction, the model detects anomalies by evaluating the prediction loss associated with masked tokens. Experimental results indicated that LAnoBERT outperformed previous unsupervised models on standard datasets including HDFS and BGL. Meanwhile, supervised learning methods have received growing interest. Guan et al. [2] proposed LogLLM, a two-stage pipeline leveraging BERT for log embeddings and a LLaMA-based decoder classifying log entries. With semantic representations aligned between the stages, LogLLM significantly improved detection accuracy in complex log scenarios. Likewise, Zhang et al. [11] introduced LogFiT, a framework that is designed to fine-tune transformer models directly on raw, unstructured log sequences, effectively managing variability and noise without relying on fixed log templates. Apart from IT-focused logs, transformer models have shown potential in other domains. For example, Liu and Buford [4] used DistilBERT to detect malicious activity in Unix shell command sequences, achieving strong performance in supervised as well as unsupervised learning. In a nutshell, these developments demonstrate that Large Language Models (LLMs) are capable of handling structured technical information beyond conventional linguistic data, establishing a basis for use in ICS environments. Building on this foundation, our study adapts a pre-trained LLM—BERT—for ICS log analysis, enabling context-sensitive anomaly detection without dependence on rule-based parsing or manual feature engineering. Unlike prior efforts focused on IT environments or semi-structured logs, our framework is designed for structured, multivariate ICS telemetry, targetting the specific security needs of cyber-physical systems.

III. Methodology

This section details the architecture of our BERT-driven anomaly detection framework tailored for logs generated by Industrial Control Systems (ICS). We present the preprocessing strategies used to convert structured ICS data into a sequential format compatible with language model training, and subsequently describe the model architecture along with the baseline methods used for comparison. Figure 1 presents the complete architecture, from data ingestion to anomaly classification.

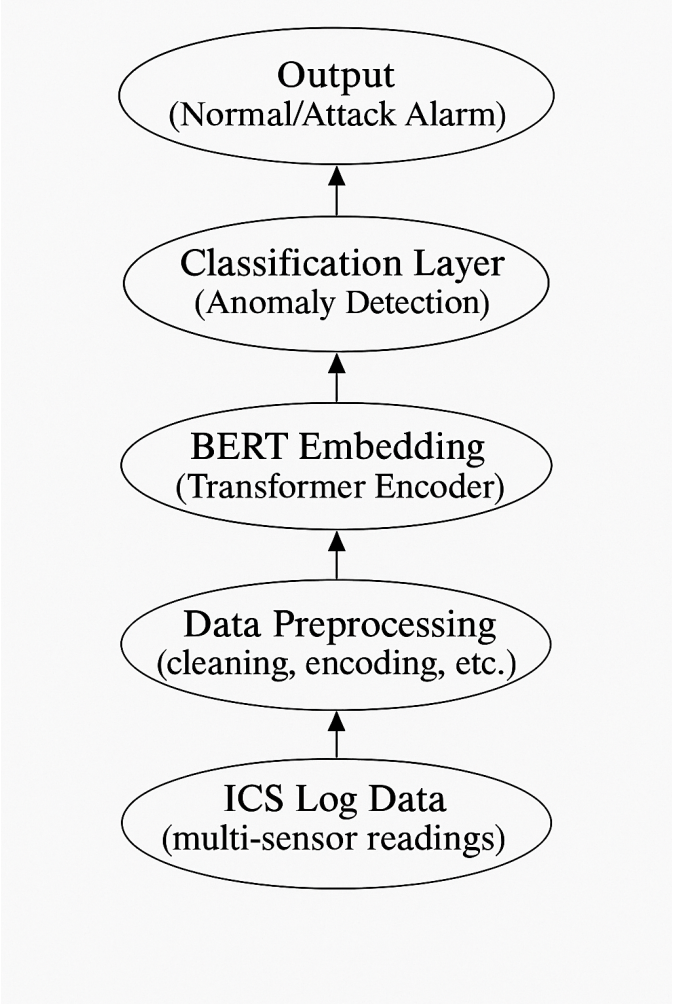

Fig. 1. Architectural Overview of the Proposed Framework.

A. Data Preprocessing and Encoding Structured logs gathered from an ICS testbed serve as the foundation for the

analysis conducted in this study, reflecting both system state variables and network-level control interactions. Each record encapsulates attributes such as device identifiers, function codes, control modes, sensor metrics, and operational timing details. These logs were captured under various operational scenarios, including multiple simulated cyber-attacks targeting different components of the control system.

To address the inherent class imbalance—where anomalous instances are rare compared to regular operations—we employed stratified sampling and oversampling techniques during training to ensure adequate exposure to attack-related patterns. The data set was divided into 80% training data and 20% test data, ensuring that all the simulated attack categories were represented proportionally in both sets.

B. Token Sequence Construction To align the ICS log format with transformer-based language models, each log entry was transformed into a serialized textual token sequence. This process involved arranging key-value pairs in a fixed, logical order, where both field names and their corresponding values were encoded as discrete tokens. The structured serialization preserved both positional and semantic relationships between fields, enabling the model to learn contextual dependencies across various operational parameters. Traditional methods rely on numerical encodings or manually extracted features; however, this representation leverages the inherent structure of ICS logs, minimizing dependence on domain-specific preprocessing steps. Placeholder tokens were introduced to represent missing or undefined values, ensuring input uniformity across all records.

C. Data Cleaning and Normalization To ensure consistent generalization across varied input conditions and enhance the model's robustness, several preprocessing steps were applied. All missing or inapplicable entries were replaced with designated placeholder tokens to maintain structural consistency. Sensor values with continuous ranges were normalized to a fixed scale prior to string token conversion. Additionally, any fields that exhibited constant values throughout the dataset were excluded to remove redundancy and minimize noise. The final tokenized representations, each depicting an ICS system state at a given moment, were paired with corresponding anomaly labels for supervised fine-tuning of the language model. This approach supports deployment across diverse ICS configurations while preserving critical system behavior characteristics.

D. Model Architecture and Anomaly Classification The proposed system utilizes BERT, a lightweight transformer-based language model, as its foundational component. Through self-attention mechanisms, the transformer encoder captures contextual dependencies across tokenized fields. The pre-trained BERT model—initially trained on large-scale general text corpora—is fine-tuned on labeled ICS log sequences for domain-specific adaptation. A classification head is appended to the model, utilizing the contextual embedding associated with a designated special token to generate binary predictions via a softmax layer. This architecture enables the model to differentiate between normal and anomalous operating states based on learned contextual patterns. After training, the system is capable of evaluating new ICS logs and classifying them as either routine activity or potential anomalies. The detection decision is based on the output probability. This ensures reliable identification of cyber-attack patterns in real time.

IV. Experimental Setup

A. Datasets

The datasets used in this study include publicly available multivariate time-series logs from the SWaT (Secure Water Treatment) testbed [22] and selected scenarios from the Morris ICS dataset repository [21]. These logs capture system behavior under both normal and simulated attack conditions, including anomalies such as address scanning, function code manipulation, parameter tampering, and command injection. Due to significant class imbalance—where anomalous events account for less than 1% of the data—stratified sampling and class-weighting techniques were applied during training to ensure representative learning. Each model was trained using 80% of the data and evaluated on the remaining 20%. For reproducibility, preprocessing was standardized: numeric fields were normalized, non-informative fields excluded, and logs were tokenized into sequences compatible with NLP-based models.

B. Implementation and Environment The anomaly detection framework was implemented using the Hugging Face Transformers library, utilizing the BERT base uncased model as its foundation. Fine-tuning was carried out on a workstation equipped with an NVIDIA RTX 3080 GPU and an Intel Core i7 processor. The model was trained on preprocessed ICS log sequences using a maximum sequence length appropriate for the input format, with most samples fitting well within this limit. Standard binary classification techniques were used along with class weighting to handle data imbalance. The training process was completed efficiently, demonstrating the feasibility of using lightweight transformer models like BERT for ICS anomaly detection in a practical computing environment.

C. Evaluation Metrics Evaluation Methodology: After training, we evaluated each model on the unseen test set (20% of data). We compute the following metrics: Accuracy: Overall fraction of correctly classified logs. Precision (for the attack class): Among all logs predicted as attacks, how many were true attacks. Recall (Detection Rate or True Positive Rate): Among all true attacks, how many did the model detect. F1-Score for the attack class: Represents the unified score derived by computing the harmonic mean between precision and recall values, offering a comprehensive metric to assess the model's overall performance in detecting anomalies.

D. Performance Metrics and Results Table 1 provides a comparative analysis of the proposed BERT-based anomaly detection framework against LSTM and TF-IDF combined with SVM baselines, evaluated on the test dataset.

Figure 2 illustrates the performance comparison between the proposed BERT-based model and two baseline methods:

TABLE I
PERFORMANCE ANALYSIS OF MODELS FOR ANOMALY DETECTION

Model	Accuracy	Precision	Recall	F1-Score
BERT (Proposed)	90.5%	0.84	0.81	**0.825**
LSTM (Baseline)	87.3%	0.80	0.76	0.78
TF-IDF + SVM	81.3%	0.62	0.48	0.54

LSTM and TF-IDF + SVM. While BERT outperformed all baseline models in terms of performance metrics, the margins over LSTM were moderate. This can be attributed to BERT's ability to learn contextual relationships within structured logs, offering better generalization in detecting subtle anomalies. The LSTM model, although simpler, performed reasonably well due to its temporal modeling strengths. The TF-IDF + SVM baseline, which lacks sequence awareness, showed lower F1-scores, particularly struggling with recall. Overall, the comparison suggests that incorporating pretrained language models offers measurable benefits, but baseline models remain competitive depending on the use case and constraints.

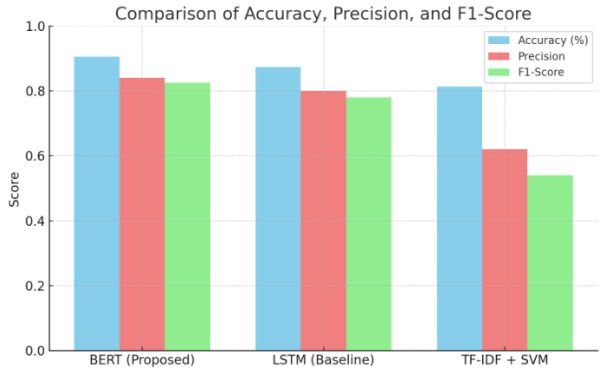

Fig. 2. Comparison of Accuracy, Precision, and F1-Score across BERT, LSTM, and TF-IDF + SVM models.

E. Discussion and comparison with existing frameworks

Most existing ICS anomaly detection frameworks rely on domain-specific methods such as physical invariant modeling, statistical thresholds, or traditional machine learning algorithms. While effective in controlled scenarios, these approaches often lack adaptability to diverse environments and require extensive domain expertise for deployment. Recent advances have introduced deep learning-based models, such as LSTMs and autoencoders, which improve temporal modeling but still depend on structured feature engineering or require large labeled datasets. Existing frameworks such as LAnoBERT and LogLLM have illustrated the applicability of language models for log-based anomaly detection; however, their focus remains largely confined to IT systems and semi-structured log data. Our proposed framework differs by applying a pre-trained language model—BERT—to fully structured ICS logs, leveraging its contextual understanding to capture nuanced inter-field dependencies without explicit feature extraction or protocol-specific rules. Compared to traditional methods, the framework provides improved generalizability, lower false positives, and reduced setup complexity. While the current results are promising, the approach may still face challenges in handling highly dynamic ICS environments or completely unseen attack types. Nonetheless, its ability to operate with minimal customization makes it a practical step toward scalable and intelligent ICS security monitoring.

V. CONCLUSION

By transforming structured ICS logs into tokenized sequences, the model effectively captured contextual relationships across multiple fields, enabling it to distinguish between normal operations and cyber-attacks with minimal reliance on handcrafted features or protocol-specific rules. The approach demonstrated consistent performance improvements over baseline models, particularly in precision, recall, and false positive reduction, across several public ICS datasets.

The ability to fine-tune a pre-trained transformer model on domain-specific log data, without requiring deep system knowledge or feature engineering, Offers a robust and adaptable approach for monitoring ICS security in real-time with high efficiency. Our results suggest that even LLM like BERT can provide meaningful insights into ICS behavior when properly adapted, and that language models can generalize beyond natural language to structured sensor and control data—offering new opportunities in cyber-physical security analytics.

VI. FUTURE WORK

Future research may focus on extending this framework to unsupervised or semi-supervised settings, which are critical for real-world deployments where labeled attack data is scarce or incomplete. Enhancing model explainability through attention visualization or hybrid integration with rule-based systems could improve trust and interpretability. Furthermore, domain-adaptive pretraining using ICS manuals, logs, and operational data may strengthen performance while reducing the need for extensive labeled datasets. Finally, deployment in live ICS environments will be essential to evaluate robustness against operational variability, concept drift, and adversarial behaviors—ensuring long-term reliability in dynamic industrial settings.

REFERENCES

[1] Abshari, D., Fu, C. and Sridhar, M., 2024. LLM-assisted Physical Invariant Extraction for Cyber-Physical Systems Anomaly Detection. arXiv preprint arXiv:2411.10918.
[2] Guan, W., Cao, J., Qian, S., Gao, J. and Ouyang, C., 2024. LogLLM: Log-based Anomaly Detection Using Large Language Models. arXiv preprint arXiv:2411.08561.
[3] Lee, Y., Kim, J. and Kang, P., 2023. Lanobert: System log anomaly detection based on bert masked language model. Applied Soft Computing, 146, p.110689.
[4] Liu, Z. and Buford, J., 2023. Anomaly detection of command shell sessions based on distilbert: Unsupervised and supervised approaches. arXiv preprint arXiv:2310.13247.
[5] Liu, J., Guogang, W.A.N.G., Zong, X., Bowei, N.I.N.G. and He, K., 2025. EfficientTransformer: A Dynamic Anomaly Detection Model for Industrial Control Networks. IEEE Access.
[6] Dehlaghi-Ghadim, A., Moghadam, M.H., Balador, A. and Hansson, H., 2023. Anomaly detection dataset for industrial control systems. IEEE Access, 11, pp.107982-107996.

[7] Ortega-Fernandez, I., Sestelo, M., Burguillo, J.C. and Piñón-Blanco, C., 2024. Network intrusion detection system for DDoS attacks in ICS using deep autoencoders. Wireless Networks, 30(6), pp.5059-5075.
[8] Fung, C., Zeng, E. and Bauer, L., 2024. Attributions for ML-based ICS anomaly detection: From theory to practice. In Proc. 31st Netw. Distrib. Syst. Secur. Symp.
[9] Gaggero, G.B., Armellin, A., Portomauro, G. and Marchese, M., 2024. Industrial control system-anomaly detection dataset (ICS-ADD) for cyber-physical security monitoring in smart industry environments. IEEE Access.
[10] Zhu, Q., Ding, Y., Jiang, J. and Yang, S.H., 2025. Anomaly detection using invariant rules in Industrial Control Systems. Control Engineering Practice, 154, p.106164.
[11] Almodovar, C., Sabrina, F., Karimi, S. and Azad, S., 2024. LogFiT: Log anomaly detection using fine-tuned language models. IEEE Transactions on Network and Service Management, 21(2), pp.1715-1723.
[12] Kim, S., Jo, W., Kim, H., Choi, S., Jung, D.I., Choi, H. and Shon, T., 2024. Two-Phase Industrial Control System Anomaly Detection Using Communication Patterns and Deep Learning. Electronics, 13(8), p.1520.
[13] Hoang, N.X., Hoang, N.V., Du, N.H., Huong, T.T. and Tran, K.P., 2022. Explainable anomaly detection for industrial control system cybersecurity. IFAC-PapersOnLine, 55(10), pp.1183-1188.
[14] Shang, W., Cui, J., Wan, M., An, P. and Zeng, P., 2016, November. Modbus communication behavior modeling and SVM intrusion detection method. In Proceedings of the 6th international conference on communication and network security (pp. 80-85).
[15] Guo, H., Yuan, S. and Wu, X., 2021, July. Logbert: Log anomaly detection via bert. In 2021 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.
[16] Ferrag, M.A., Ndhlovu, M., Tihanyi, N., Cordeiro, L.C., Debbah, M., Lestable, T. and Thandi, N.S., 2024. Revolutionizing cyber threat detection with large language models: A privacy-preserving bert-based lightweight model for iot/iiot devices. IEEe Access, 12, pp.23733-23750.
[17] Aghaei, E., Niu, X., Shadid, W. and Al-Shaer, E., 2022, October. Securebert: A domain-specific language model for cybersecurity. In International Conference on Security and Privacy in Communication Systems (pp. 39-56). Cham: Springer Nature Switzerland.
[18] Jaradat, S., Komol, M.M., Elhenawy, M. and Dong, N., 2024. Cyberattack detection on SWaT plant industrial control systems using machine learning. Artif. Intell. Auto. Syst.
[19] Madupati, B., 2022. Machine Learning for Cybersecurity in Industrial Control Systems (ICS). Available at SSRN 5076696.
[20] Lamberts, O., Wolsing, K., Wagner, E., Pennekamp, J., Bauer, J., Wehrle, K. and Henze, M., 2023. SoK: Evaluations in industrial intrusion detection research. arXiv preprint arXiv:2311.02929.
[21] Morris, T. and Gao, W., 2014. Industrial control system traffic data sets for intrusion detection research. In Critical Infrastructure Protection VIII: 8th IFIP WG 11.10 International Conference, ICCIP 2014, Arlington, VA, USA, March 17-19, 2014, Revised Selected Papers 8 (pp. 65-78). Springer Berlin Heidelberg.
[22] Goh, J., Adepu, S., Junejo, K.N. and Mathur, A., 2017. A dataset to support research in the design of secure water treatment systems. In Critical Information Infrastructures Security: 11th International Conference, CRITIS 2016, Paris, France, October 10–12, 2016, Revised Selected Papers 11 (pp. 88-99). Springer International Publishing.
[23] López-Morales, Efrén, et al. ”SoK: Security of Programmable Logic Controllers.” 33rd USENIX Security Symposium (USENIX Security 24). 2024.