PureChain-Based Zero-Knowledge Proofs for Verifiable Machine Learning in Industrial IoT

George Chidera Akor [1], Love Allen Chijioke Ahakonye [2], Jae Min Lee [1], Dong-Seong Kim [1] [*]
[1] IT-Convergence Engineering, *Kumoh National Institute of Technology*, Gumi, South Korea
[2] ICT Convergence Research Center, *Kumoh National Institute of Technology*, Gumi, South Korea
[*] NSLab Co. Ltd., Gumi, South Korea, *Kumoh National Institute of Technology*, Gumi, South Korea
(georgeakor, loveahakonye, ljmpaul, dskim@kumoh.ac.kr)

***Abstract*—Industrial IoT relies on edge machine learning for anomaly detection but often lacks cryptographic guarantees, making alerts vulnerable to manipulation. Existing methods face challenges in auditability and are costly for on-chain verification. We propose a pipeline combining a sparse autoencoder with the general purpose veriable machine learning library, known as EZKL. The EZKL zero-knowledge proofs are verified on the PureChain, providing verifiable and tamper-evident alerts. On the Edge-IIoTset dataset, the system achieves ROC-AUC 0.971 and PR-AUC 0.753. Proof generation takes 69-70 seconds per sample, with on-chain verification stable at 0.80 seconds and minimal calldata. Future work will focus on hardware acceleration, model compression, and scalable proof aggregation.**

***Index Terms*—blockchain-as-a-service, Industrial IoT, intrusion detection, PureChain, verifiable inference, Zero-knowledge machine learning**

I. Introduction

Edge machine learning is increasingly used to protect operational technology networks [1]. However, inference often runs on untrusted gateways where alerts may be forged or suppressed [2]. We close this integrity gap by binding anomaly detections to zero-knowledge proofs that can be verified on-chain with near-constant cost. The approach combines a lightweight autoencoder with zero-knowledge proof generation and on-chain verification on the PureChain network, yielding an immutable and portable audit trail for security operations.

The idea builds on work at the intersection of blockchain and IoT security [3] and on industrial directions such as blockchain-as-a-service and pure-chain stacks [4], [5]. Our contribution is an executable pipeline with end-to-end measurements of model skill, proving cost, calldata footprint, and verification latency.

II. Methodology

We utilize the Edge-IIoTset dataset [6] for training a sparse denoising autoencoder with a 60! →!32! →!12! →!32! →!60 architecture. Training incorporates Gaussian noise, dropout, and mild ℓ_1/ℓ_2 regularization, with the reconstruction error measured by mean absolute error. Evaluation metrics include ROC-AUC, PR-AUC, and confusion matrices at the optimal threshold, selected based on a validation sweep of reconstruction errors; the threshold for our run was 6.505×10^{-2}.

For zero-knowledge machine learning, the trained Keras model is converted to an ONNX inference graph, compiled into a proving circuit using EZKL [7]. The proving and verification keys are generated, and a Solidity verifier is created. Proofs and witnesses are produced via the EZKL Python API [8] and encoded to EVM calldata. On-chain verification is performed on the PureChain [5] by deploying the verifier and using `eth_call` for gas-free measurements. Extra data handling is applied if required by PoA middleware. Fig 1 illustrates this approach and the entire pipeline, from preprocessing to proof generation and verification, runs in the supplied Colab notebook[1].

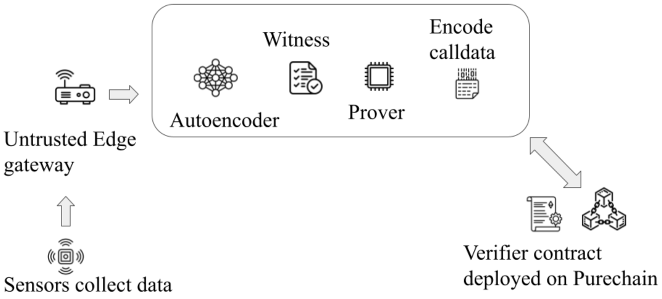

Fig. 1. System diagram. Sensors stream data to an untrusted edge gateway that runs a sparse autoencoder. The anomaly score and features generate a witness. EZKL then produces a proof and encodes calldata, which is sent to a verifier contract deployed on PureChain. Verification yields a tamper-evident record while raw sensor data remain off chain.

III. Results and Discussion

The detector achieves an ROC-AUC of 0.971 and a PR-AUC of 0.753. At the optimal F_1 threshold ($\tau = 6.505 \times 10^{-2}$), precision is 0.746, recall is 1.000, F_1 is 0.854, and the false-positive rate is 0.0589. Zero-knowledge measurements indicate that proving dominates runtime, with a mean proving time of approximately 69–70 seconds per sample across ten samples on a Colab virtual machine (Fig. 2). Witness generation and encoding contribute minimally, as shown in Fig. 3. On-chain verification via `eth_call` remains stable around 0.80 seconds (Fig. 4). Calldata size averages at 7.65KB with minimal variation (Fig. 5). Table I summarizes the results, noting a 100% verification success rate across ten runs.

The measurements indicate a clear division of labor: the chain ensures efficient, predictable verification and compact

[1] Colab link

audit records, while the prover handles the majority of the computational cost. This setup benefits systems that batch or stream alerts, as proofs can be generated off-path and verified asynchronously. The main limitation is prover latency, which can be mitigated through hardware acceleration, optimized activations, and model compression.

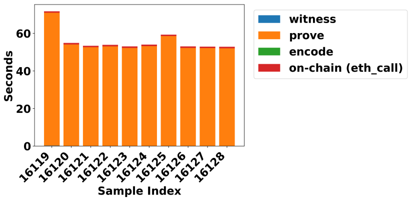

Fig. 2. Per-sample breakdown: witness, prove, encode and on-chain (`eth_call`).

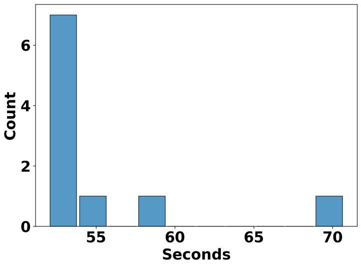

Fig. 3. Histogram of proving times across samples.

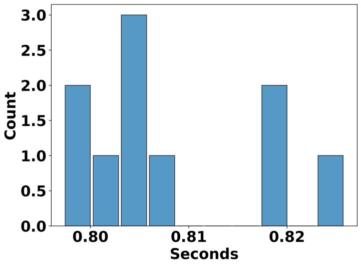

Fig. 4. Histogram of on-chain (`eth_call`) verification times.

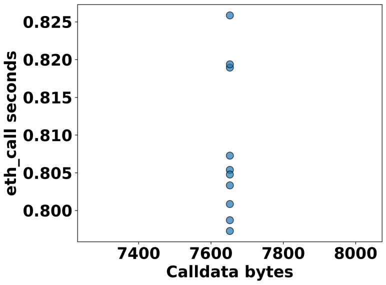

Fig. 5. Calldata size vs. on-chain verification latency.

TABLE I
ZKML AND ON-CHAIN METRICS (10 SAMPLES)

Metric	Mean	P95
Witness (s)	0.06	–
Prove (s)	55.02	65.19
Encode (s)	0.00	–
On-chain eth_call (s)	0.808	0.823
End-to-end (s)	56.00	–
Calldata (bytes)	7652	–
On-chain TRUE rate	100.0%	–

IV. Conclusion

This study demonstrate a complete and reproducible pipeline for verifiable IIoT anomaly detection in which each alert can be accompanied by a succinct proof and an on-chain verification outcome. The system delivers a tamper-evident audit trail with constant-time verification and a small calldata footprint. Future work will focus on accelerating the prover, exploring sequence models tailored to lookup-friendly operations and aggregating proofs so that rolling audit logs remain efficient at scale.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2025-RS-2020-II201612, 33%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33%) and by the MSIT, Korea, under the ITRC support program(IITP-2025-RS-2024-00438430, 34%).

References

[1] T. Yang, X. Jiang, W. Li, P. Liu, J. Wang, W. Hao *et al.*, "Cloud–edge collaborative data anomaly detection in industrial sensor networks," *PLOS ONE*, vol. 20, no. 6, p. e0324543, 2025.

[2] P. Mahadevappa, R. Al-amri, G. Alkawsi, A. A. Alkahtani, M. F. Alghenaim, and M. Alsamman, "Analyzing threats and attacks in edge data analytics within iot environments," *IoT*, vol. 5, no. 1, pp. 123–154, 2024.

[3] L. A. C. Ahakonye, C. I. Nwakanma, and D.-S. Kim, "Tides of blockchain in iot cybersecurity," *Sensors*, vol. 24, no. 10, p. 3111, 2024.

[4] D.-S. Kim, E. A. Tuli, I. S. Igboanusi, and M. M. H. Somrat, "Blockchain-as-a-service: A pure chain approach," *Blockchain: Research and Applications*, p. 100397, 2025.

[5] L. A. C. Ahakonye, C. I. Nwakanma, J. M. Lee, and D. S. Kim, "Purechain-enhanced federated learning for dynamic fault tolerance and attack detection in distributed systems," *High-Confidence Computing*, p. 100354, 2025.

[6] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-iiotset: A new comprehensive realistic cyber security dataset of iot and iiot applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40 281–40 306, 2022.

[7] *The EZKL System — Official Documentation*, Zkonduit Inc., 2025, accessed 2025-10-20. [Online]. Available: https://docs.ezkl.xyz/

[8] *ezkl Python bindings — Documentation*, Zkonduit Inc., 2025, accessed 2025-10-20. [Online]. Available: https://pythonbindings.ezkl.xyz/en/latest/