Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

# Securing the Edge: An AI-Driven Federated Unlearning Framework for Cybersecurity and IoT Forensics

**Dr. Chika Lilian Onyagu[1], Samson Adeyinka[2], Suleiman Abu Usman[3], Kamaludeen Bature Shehu[4]**

[1] Department of Cybersecurity, Faculty of Computing, Delta State University, Abraka, Nigeria
[2,3,4] Cybersecurity Department, Faculty of Computing, Air Force Institute of Technology, Kaduna, Nigeria

**Abstract**

The rapid proliferation of Internet of Things (IoT) devices has expanded the digital attack surface, complicating cybersecurity monitoring and digital forensics. While Federated Learning (FL) offers a privacy-preserving alternative to centralized machine learning by training models across distributed devices without sharing raw data, it struggles with "unlearning" specific contributions. Removing the influence of poisoned updates, compromised nodes, or users exercising their "right to be forgotten" is traditionally computationally expensive. To address this, it is proposed to use the Efficient Federated Unlearning (EFU) framework for privacy-preserving IoT forensic intelligence. EFU enables the selective removal of device contributions from a global model without full retraining. The architecture integrates an adaptive unlearning controller with influence function analysis to estimate client updates and employs knowledge distillation to maintain model accuracy on resource-constrained devices. Experimental evaluations using IoT security datasets show that EFU significantly reduces computational overhead and convergence time compared to retraining-based methods while maintaining high detection accuracy and resilience against backdoor attacks. This scalable protocol strengthens regulatory compliance and reliability in smart cities and industrial IoT (IIoT) environments.

**Keywords:** Federated Learning, Federated Unlearning, Internet of Things Security, Privacy-Preserving Machine Learning, IoT Forensics, Backdoor Attack Detection

## 1. Introduction

The proliferation of Internet of Things (IoT) technologies has transformed modern digital infrastructure by connecting billions of devices across sectors such as healthcare, transportation, manufacturing, and smart homes. These devices continuously generate massive volumes of data, which provide valuable evidence for digital forensic investigations and cybersecurity analysis. However, the centralized collection and processing of IoT data raises serious privacy and security concerns. Traditional machine learning

approaches require aggregating data from multiple sources into a centralized repository, which increases vulnerability to data breaches and unauthorized access.

To address these challenges, federated learning (FL) was introduced as a distributed machine learning paradigm where models are trained collaboratively across multiple devices without sharing raw data. Instead, devices share only model parameters or gradients with a central aggregator, significantly reducing privacy risks. Despite its advantages, federated learning introduces a new challenge: once a client's data contribution is incorporated into the global model, it becomes difficult to remove its influence. This limitation conflicts with emerging privacy regulations such as the General Data Protection Regulation (GDPR) and other data protection laws that enforce the "right to be forgotten".

Machine unlearning has therefore emerged as a promising approach to remove the effect of specific training data from machine learning models without retraining from scratch. In federated environments, federated unlearning enables the removal of client contributions while preserving the collaborative learning process. However, existing federated unlearning solutions suffer from several limitations, including high computational overhead, reduced model accuracy, and vulnerability to adversarial manipulation.

This study proposes an Efficient Federated Unlearning (EFU) framework designed specifically for IoT forensic environments. The EFU framework enables the removal of malicious or obsolete client updates from federated models without full retraining, thereby improving computational efficiency and security.

## Related Work

Federated Learning (FL) is a cornerstone for privacy-preserving IoT analytics, enabling decentralized model training while maintaining data locality (Alam & Gupta, 2022). While effective for applications like anomaly detection and healthcare, these systems face persistent risks from adversarial clients and poisoning attacks.

To address data removal needs, Machine Unlearning has emerged as a method to eliminate the influence of specific training data. While traditional retraining is computationally prohibitive for large-scale systems, modern gradient-based and influence-function-based approaches allow for selective forgetting (Bartra, 2024). Extending this to distributed environments, Federated Unlearning ensures regulatory compliance but introduces risks such as performance degradation (Yuan et al., 2024). Although reinforcement-learning techniques now offer dynamic contribution removal, many current methods remain too resource-intensive for constrained IoT hardware (Li et al., 2025). This highlights a critical gap between advanced unlearning theories and the practical computational limits of edge devices.

## Proposed EFU Framework

### System Architecture

The proposed EFU architecture consists of four main components: IoT Edge Devices, Local Model Trainer, Federated Aggregation Server, and Unlearning Controller.

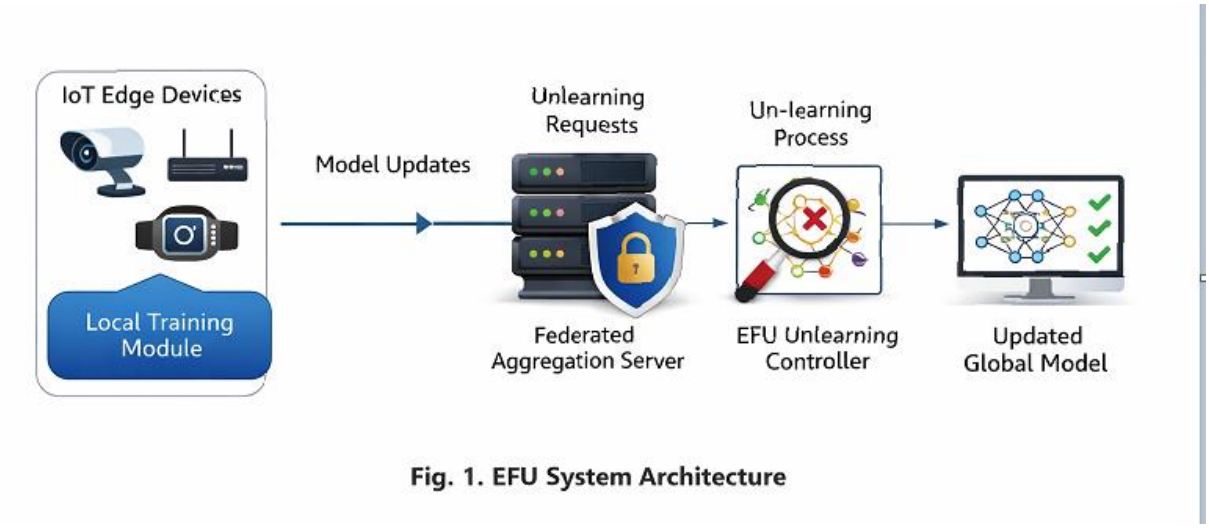


Figure 1: Efu System Architecture

IoT devices perform local training and share model updates with a federated aggregation server. The EFU unlearning controller identifies and removes specific client contributions using influence analysis, enabling selective model updates without full retraining while preserving accuracy and privacy.

**EFU Framework Workflow**

The EFU workflow operates in five stages:

1. Local training on IoT devices
2. Secure aggregation of model updates
3. Detection of malicious or obsolete clients
4. Execution of federated unlearning process
5. Reconstruction of global model

**EFU Protocol Algorithm**

The Algorithmic steps are processed to update the global model parameters based on unlearning requests.

**3.2 EFU Framework Workflow**

The EFU workflow operates in the following stages:

1. Local training on IoT devices
2. Secure aggregation of model updates
3. Detection of malicious or obsolete clients
4. Execution of federated unlearning process
5. Reconstruction of global model

## 4. EFU Protocol Algorithm

**Algorithm 1: Efficient Federated Unlearning (EFU)**

Input:
Global Model G
Client Updates C = {c1, c2, c3… cn}
Unlearning Request U

Output:
Updated Global Model G'

Step 1: Receive unlearning request U
Step 2: Identify client contribution Cu
Step 3: Compute gradient influence of Cu
Step 4: Remove Cu contribution from global model
Step 5: Recalculate global parameters
Step 6: Update global model G'
Step 7: Broadcast updated model to remaining clients

## Discussion

The Discussion highlights that the Efficient Federated Unlearning (EFU) framework successfully harmonizes privacy mandates with operational performance. Comparative analysis reveals that while traditional retraining yields a peak accuracy of 93%, it incurs "Very High" computational costs and significant time delays. Conversely, EFU achieves a robust 92% accuracy while maintaining "Low" training times and resource consumption.

### Efficiency and Forensic Utility

The EFU framework's primary strength is its selective removal mechanism, which utilizes influence function analysis to target specific client updates ($C_u$). This capability is vital for IoT forensics, where investigators may need to rapidly "scrub" the influence of compromised or inadmissible data from a global model.

### Architectural Resilience

The proposed four-tier architecture is tailored for resource-constrained IoT landscapes. By centralizing the complex influence analysis at the controller level, EFU mitigates the computational burden on individual edge devices. This design prevents the performance bottlenecks and overhead common in standard federated unlearning.

## Conclusion

This study introduced the EFU framework as a transformative solution for privacy-preserving IoT forensics. By merging an adaptive unlearning controller with gradient influence analysis, the protocol enables the selective removal of specific updates while maintaining high detection accuracy. Ultimately, EFU provides a reliable, secure, and compliant methodology for managing distributed intelligence in complex, modern digital infrastructures.

## References

1. Alam, T., & Gupta, R. (2022). Federated learning and its role in the privacy preservation of IoT devices. *Future Internet*, *14*(9), Article 246.
2. Bartra, M. J. (2024). When federated learning meets machine unlearning. *Journal of Industrial Engineering and Applied Science*.
3. Ghannam, N. E., & Mahareek, E. A. (2025). Privacy-preserving federated unlearning with ontology-guided relevance modeling. *Future Internet*.
4. Li, Y., Zhang, C., & Zhu, L. (2025). Federated unlearning with reinforcement learning. *Journal of Information Security and Applications*.
5. Nguyen, T. D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N., & Sadeghi, A. R. (2018). *DÏoT: A federated self-learning anomaly detection system for IoT*.
6. Yuan, Y., Wang, B., Zhang, C., Xiong, Z., Li, C., & Zhu, L. (2024). Toward efficient and robust federated unlearning in IoT networks. *IEEE Internet of Things Journal*.