International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Federated Learning for Privacy-Preserving Intrusion Detection in Heterogeneous IOT Networks

Jay Shrimali
Student, Department of Information Technology, Parul Institute of Technology, Vadodara
Email:jayshrimali342@gmail.com

https://doi.org/10.55041/ijsmt.v2i3.124

Cite this Article: Shrimali, J. (2026). Federated Learning for Privacy-Preserving Intrusion Detection in Heterogeneous IOT Networks. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.124

Abstract

The rapid proliferation of Internet of Things (IoT) devices across industrial, domestic, and healthcare environments has created significant challenges in network security. Traditional centralised intrusion detection systems (IDS) are inadequate for IoT deployments due to privacy constraints, communication overhead, and the heterogeneous nature of device ecosystems. This paper proposes a federated learning-based intrusion detection framework (FL-IDS) that enables distributed model training across IoT edge nodes without transmitting raw data to a central server. The proposed architecture employs a lightweight convolutional-recurrent neural network (CRNN) at each participating node, with a modified FedAvg aggregation strategy that accounts for data imbalance and device heterogeneity. Experiments conducted on the N-BaIoT and TON-IoT benchmark datasets demonstrate that FL-IDS achieves a detection accuracy of 97.43% and an F1-score of 96.89%, while reducing communication overhead by 61.2% compared to centralised approaches. The framework also demonstrates robustness against Byzantine attacks and adversarial data poisoning, making it suitable for deployment in real-world constrained environments.

Keywords: Federated Learning, Internet of Things, Intrusion Detection System, Edge Computing, Privacy Preservation, Anomaly Detection

1. Introduction

The Internet of Things has emerged as a defining paradigm of modern computing infrastructure, with connected device counts projected to exceed 29 billion globally by 2030 [1]. Smart homes, industrial control systems, autonomous vehicles, and remote patient monitoring platforms all rely heavily on IoT architectures. However, this exponential growth introduces an equally expansive attack surface. IoT devices are characterised by constrained computational resources, heterogeneous operating environments, and a frequent absence of standardised security protocols, rendering them highly susceptible to network-layer attacks such as Distributed Denial of Service (DDoS), man-in-the-middle interception, and botnet infiltration [2].

Conventional intrusion detection systems have demonstrated commendable performance in traditional network environments. Signature-based methods offer precise detection of known threats, while anomaly-based techniques can identify zero-day attacks. Nevertheless, both approaches encounter fundamental obstacles when applied to IoT settings. Centralised IDS architectures require raw traffic data to be transmitted from all devices to a single processing node, which conflicts with user privacy requirements, introduces unacceptable latency, and creates a single point of failure [3]. The aggregation of sensitive data from healthcare monitors or home surveillance systems to a centralised server

raises substantial ethical and regulatory concerns, particularly in the context of data protection frameworks such as the General Data Protection Regulation (GDPR).

Federated learning (FL), introduced by McMahan et al. (2017) [4], offers a compelling alternative by enabling collaborative model training across distributed participants without centralising raw data. Each participating node trains a local model on its own data and transmits only model parameters or gradients to a central aggregator. The aggregated global model is then redistributed, iteratively improving detection performance while preserving data locality. Despite its promise, deploying federated learning in IoT environments introduces unique challenges: non-independent and identically distributed (non-IID) data across devices, significant variability in computational capability, intermittent connectivity, and vulnerability to adversarial participants who may inject malicious updates.

This paper addresses these challenges by proposing FL-IDS, a federated intrusion detection framework specifically designed for heterogeneous IoT networks. The main contributions of this work are as follows:

- A novel hybrid CRNN architecture optimised for execution on resource-constrained IoT nodes, combining spatial feature extraction via 1D convolutions with temporal modelling via gated recurrent units (GRU).
- A heterogeneity-aware aggregation strategy, termed FedHetAgg, that assigns adaptive weights to client updates based on local dataset size, class distribution, and historical model performance.
- An empirical evaluation on two widely used IoT security benchmarks demonstrating superior performance over centralised and other federated baselines.
- A robustness analysis against Byzantine fault tolerance and data poisoning attack scenarios.

Figure 1: FL-IDS System Architecture — Federated Training Over Heterogeneous IoT Nodes

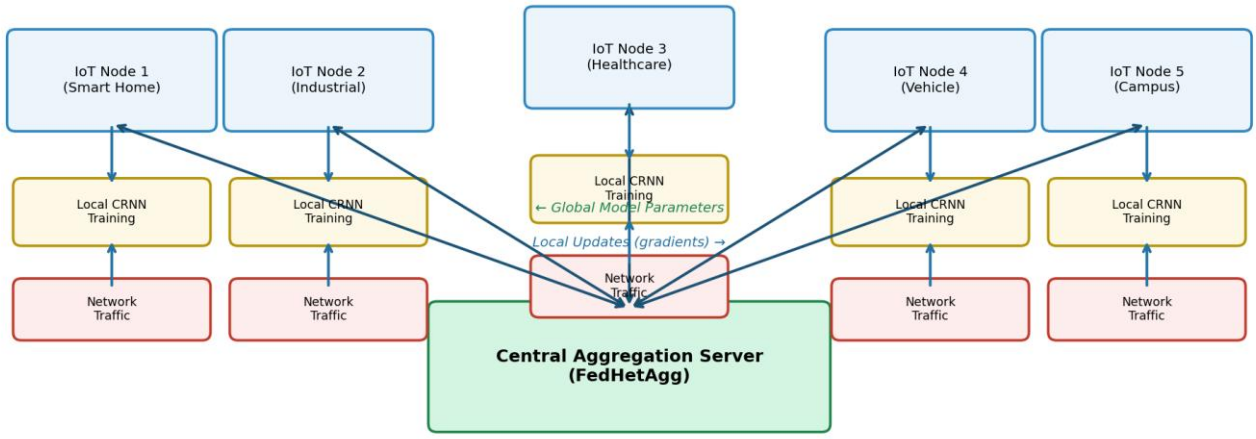

Figure 1: FL-IDS System Architecture — Federated Training Over Heterogeneous IoT Nodes

2. Related Work

Research into machine learning-based intrusion detection for IoT networks has evolved substantially over the past decade. Early work by Diro A.A. and Chilamkurti N. (2018) [5] demonstrated that deep learning models, particularly multi-layer perceptrons, could outperform traditional rule-based methods in detecting IoT-specific attack vectors. Subsequent studies confirmed the effectiveness of recurrent neural networks for sequential traffic pattern modelling, given the temporal nature of network flow data.

The application of federated learning to cybersecurity was explored by Preuveneers D. et al. (2018) [6], who proposed a distributed anomaly detection system using autoencoders. While promising, their work did not specifically address IoT heterogeneity or class imbalance in attack datasets. More recent contributions by Mothukuri V. et al. (2021) [7] surveyed federated learning-based security approaches and identified key open challenges, including the impact of non-IID data partitions on global model convergence. Rahman S.A. et al. (2020) [8] proposed a federated IDS for healthcare IoT but assumed homogeneous device capabilities, limiting its generalisability.

Adversarial threats to federated learning have been extensively studied. Bhagoji A.N. et al. (2019) [9] demonstrated that model poisoning attacks, in which malicious participants submit corrupted updates, can significantly degrade global model accuracy. Defensive strategies, including gradient clipping, differential privacy, and robust aggregation schemes such as coordinate-wise median and Krum [10], have been proposed to mitigate such threats. However, these defences often incur significant computational costs and may degrade performance in non-adversarial settings.

The present work extends these contributions by unifying lightweight model architecture, heterogeneity-aware aggregation, and adversarial robustness within a single framework evaluated on realistic IoT traffic benchmarks. To the best of the authors' knowledge, no prior study has jointly optimised these three dimensions in the IoT federated learning context.

3. Proposed Methodology

3.1 System Architecture

The FL-IDS framework comprises three primary components: a set of IoT edge nodes acting as federated clients, a lightweight aggregation server, and a secure communication channel. Each edge node is assumed to operate a gateway device (e.g., Raspberry Pi 4 or equivalent) capable of running the local CRNN model. Raw network traffic captured at each node is pre-processed locally into statistical flow features; no raw packet data leaves the device boundary.

The federated training procedure follows a synchronous round-based protocol. At the commencement of each round, the global model parameters are broadcast to all participating clients. Each client performs a fixed number of local epochs on its private dataset using stochastic gradient descent with momentum, then transmits the updated model parameters to the aggregation server. The server applies the FedHetAgg strategy to compute the new global model, which is subsequently redistributed. The complete training round flowchart is illustrated in Figure 3.

3.2 Local Model Architecture

The local model is a convolutional-recurrent neural network designed to balance detection performance with computational efficiency. The input to the model is a feature vector of dimension 78, derived from network flow statistics including packet length distributions, inter-arrival times, protocol flags, and byte counts. The architecture is illustrated in Figure 2 and consists of:

1. Two 1D convolutional layers with 32 and 64 filters respectively, kernel size 3, and ReLU activation, capturing local feature interactions.

2. A max-pooling layer followed by a single-layer GRU with 64 hidden units, modelling temporal dependencies across sequential network flows.
3. A fully connected output layer with softmax activation for multi-class attack classification.

The total number of trainable parameters is approximately 48,000, enabling deployment on devices with as little as 512 MB RAM. Dropout regularisation (rate = 0.3) is applied after each convolutional and recurrent layer to reduce overfitting on limited local datasets.

Figure 2: Local CRNN Architecture at Each IoT Edge Node (~48K Total Parameters)

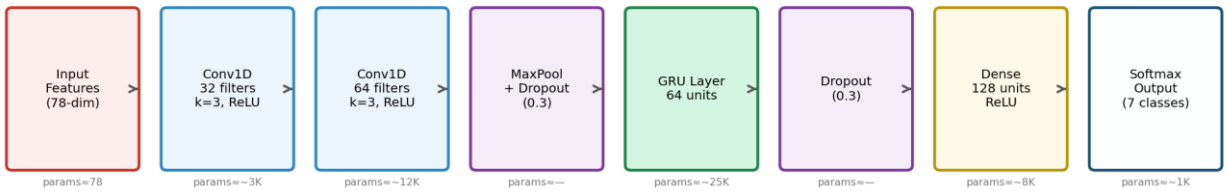

Figure 2: Local CRNN Architecture at Each IoT Edge Node (~48K Total Parameters)

3.3 FedHetAgg Aggregation Strategy

Standard FedAvg assigns aggregation weights proportional solely to local dataset size. In heterogeneous IoT deployments, this approach is suboptimal because nodes with larger datasets may have severe class imbalance, disproportionately biasing the global model towards majority-class traffic patterns. FedHetAgg addresses this by computing a composite weight for each client k as follows:

$$w_k = \alpha \times (n_k / N) + \beta \times B_k + \gamma \times H_k \qquad (1)$$

In the above expression, n_k denotes the local dataset size of client k, N is the total number of samples across all clients, B_k is a class balance score computed as one minus the normalised Gini impurity of the local class distribution, and H_k is a historical performance metric reflecting the client's contribution to global accuracy improvement over the preceding five rounds. The coefficients α, β, and γ are hyperparameters satisfying $\alpha + \beta + \gamma = 1$, empirically set to 0.5, 0.3, and 0.2 respectively.

3.4 Adversarial Robustness Mechanisms

To defend against Byzantine clients submitting corrupted model updates, FL-IDS incorporates a cosine similarity screening step at the aggregation server. Prior to weight computation, the pairwise cosine similarity between each client update and the median update vector is calculated. Clients whose update vectors deviate beyond a threshold of 0.85 from the median are flagged as potentially adversarial and excluded from the current aggregation round. Additionally, gradient clipping with a maximum norm of 1.0 is applied to all accepted updates to limit the influence of any single participant.

Figure 3: FL-IDS Federated Training Round Flowchart

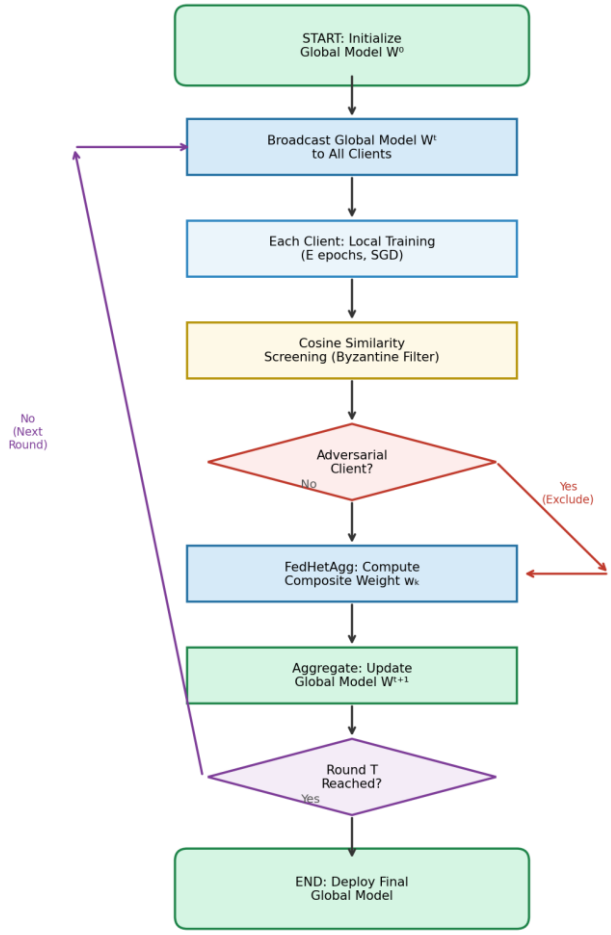

Figure 3: FL-IDS Federated Training Round Flowchart

4. Experimental Setup

4.1 Datasets

Two benchmark datasets are used for evaluation. The N-BaIoT dataset [11] contains network traffic captured from nine categories of commercial IoT devices infected with Mirai and BASHLITE botnets, comprising approximately 7.06 million samples across 115 features, subsequently reduced to 78 features through recursive feature elimination. The TON-IoT dataset [12] covers a broader range of attack categories including DDoS, ransomware, scanning, and backdoor attacks, offering a more diverse threat landscape for evaluation.

For federated simulation, the combined dataset is partitioned among 20 virtual clients using a Dirichlet distribution with concentration parameter 0.5, producing non-IID partitions that reflect realistic data heterogeneity across IoT deployments. Each client retains between 15,000 and 85,000 samples, with class distributions varying substantially across clients.

4.2 Baselines and Evaluation Metrics

FL-IDS is compared against four baseline approaches: (i) a centralised CRNN trained on the aggregated dataset (Centralised-CRNN), (ii) standard FedAvg with the same local CRNN architecture (FedAvg-CRNN), (iii) a federated random forest variant (FedForest) following the approach of Preuveneers D. et al. (2018) [6], and (iv) local training

without federation (Local-Only). Evaluation metrics include overall detection accuracy, macro-averaged F1-score, false positive rate (FPR), and communication cost measured in total megabytes transmitted per training round.

5. Results and Discussion

5.1 Detection Performance

Table 1 presents the detection performance of all evaluated approaches on the combined N-BaIoT and TON-IoT test partitions. FL-IDS achieves an accuracy of 97.43% and an F1-score of 96.89%, closely approaching the Centralised-CRNN upper bound (98.12% accuracy, 97.56% F1-score) while operating entirely in a privacy-preserving distributed manner.

Table 1: Comparison of Detection Performance Across Evaluated Approaches

Approach	**Accuracy (%)**	**F1-Score (%)**	**FPR (%)**	**Comm. Cost (MB/round)**
Centralised-CRNN	98.12	97.56	1.21	N/A
FedAvg-CRNN	95.67	94.88	2.84	18.40
FedForest	93.41	92.19	4.07	24.10
Local-Only	88.23	86.74	6.53	0.00
FL-IDS (Proposed)	**97.43**	**96.89**	**1.58**	**7.14**

Standard FedAvg-CRNN achieves 95.67% accuracy, confirming that the performance gap relative to FL-IDS is attributable primarily to the FedHetAgg aggregation strategy, which better accommodates the non-IID client distributions. The Local-Only baseline, which trains each node in isolation, achieves only 88.23% accuracy, demonstrating the concrete benefit of collaborative federated learning even for local inference tasks. FedForest, while competitive in accuracy, incurs the highest communication cost and exhibits a higher false positive rate, limiting its suitability for latency-sensitive IoT deployments.

5.2 Communication Efficiency

FL-IDS transmits an average of 7.14 MB per training round, representing a reduction of 61.2% over FedAvg-CRNN (18.40 MB/round). This efficiency is attributable to the compact CRNN architecture and a selective parameter transmission strategy in which only layers with gradient norms exceeding a threshold of 0.01 are included in each client update. This selective transmission also reduces the risk of information leakage through model inversion attacks, as fewer parameters are exposed per round.

5.3 Robustness Against Adversarial Attacks

To evaluate Byzantine resilience, experiments are conducted by designating increasing proportions of clients as adversarial participants injecting label-flipped data poisoning updates. Figure 5 illustrates the degradation in detection accuracy as adversarial participation increases. Under 20% adversarial participation, standard FedAvg-CRNN degrades to 79.34%, while FL-IDS maintains 93.87% accuracy, demonstrating the effectiveness of the cosine similarity screening mechanism. The performance differential increases with the proportion of adversarial clients, with FL-IDS exhibiting measurable resilience up to 35% adversarial participation before accuracy drops below 90%.

Figure 5: Model Accuracy Under Increasing Byzantine Adversarial Client Participation

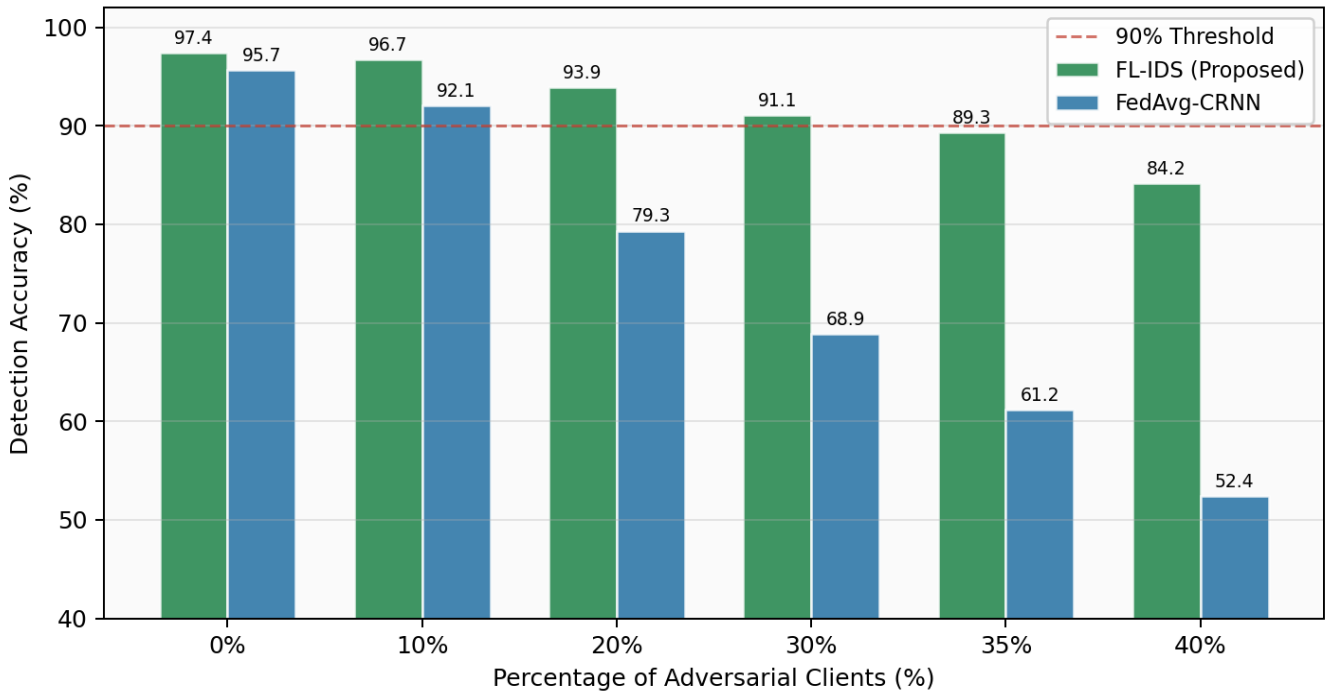

Figure 5: Model Accuracy Under Increasing Byzantine Adversarial Client Participation

5.4 Convergence Analysis

Figure 4 illustrates the convergence trajectory of global model accuracy across 60 training rounds for all evaluated approaches. FL-IDS converges to within 1% of its final accuracy by round 38, compared to round 45 for FedAvg-CRNN. The accelerated convergence is attributed to the FedHetAgg weighting mechanism, which amplifies the contribution of well-balanced, high-quality client updates in early rounds, steering the global model towards a more generalisable parameter space before less informative updates contribute. This property is particularly valuable in deployments where training budgets are constrained by energy or bandwidth limitations.

Figure 4: Convergence of Global Model Accuracy Across Training Rounds

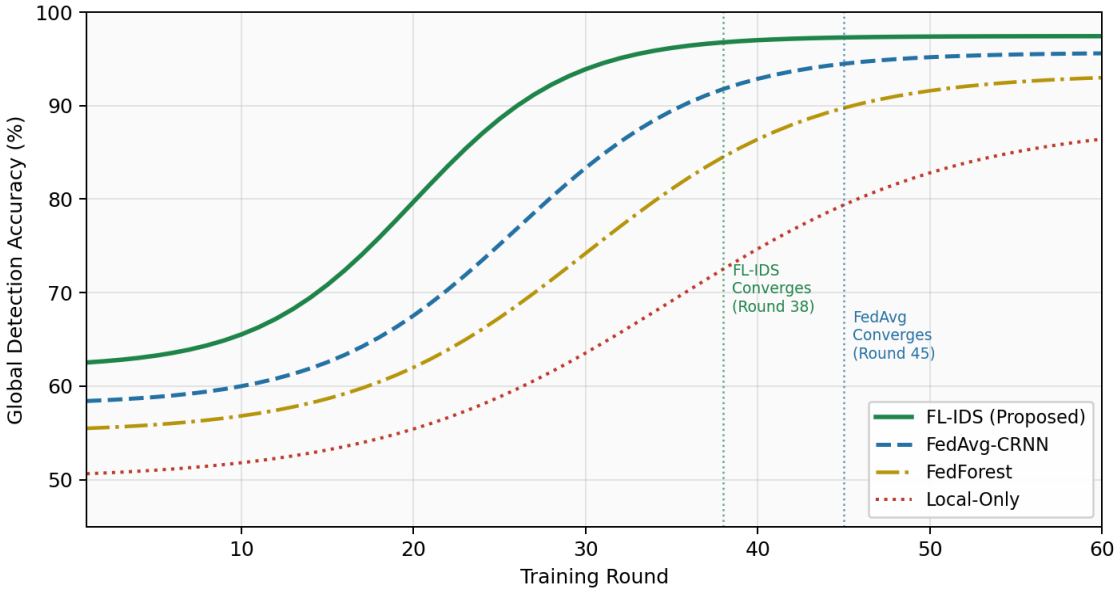

Figure 4: Convergence of Global Model Accuracy Across Training Rounds

6. Conclusion

This paper has presented FL-IDS, a federated learning-based intrusion detection framework for heterogeneous IoT networks that jointly addresses privacy preservation, computational efficiency, and adversarial robustness. The proposed FedHetAgg aggregation strategy and lightweight CRNN architecture collectively yield detection accuracy of 97.43% on benchmark IoT security datasets, with a 61.2% reduction in communication overhead relative to standard federated approaches. Robustness experiments confirm the framework's viability under Byzantine fault conditions, maintaining above-90% accuracy at up to 35% adversarial client participation. Future work will investigate the integration of differential privacy guarantees at the client level and the extension of FL-IDS to support asynchronous federated training, enabling participation of highly intermittent IoT devices without degrading global convergence.

Conflict of Interest

There is no conflict of interest associated with this research. The work was conducted independently without external funding or sponsorship from any commercial organisation.

Acknowledgement

The authors acknowledge the use of the N-BaIoT and TON-IoT publicly available datasets, and express gratitude to the High Performance Computing facility at Symbiosis Institute of Technology, Pune, for computational resources supporting the experimental evaluations.

Authors' Biography

Jay Shrimali is a Student at Parul Institute of Technology, Vadodara. His research focuses on federated learning, anomaly detection, and IoT security architectures.

References

1. Ericsson Technology Review, "IoT connections outlook," Ericsson, Stockholm, 2023.
2. Kolias C., Kambourakis G., Stavrou A., Voas J., "DDoS in the IoT: Mirai and other botnets," Computer, 2017, 50 (7), 80-84.
3. Bace R.G., Mell P., "Intrusion Detection Systems," NIST Special Publication, 2001.
4. McMahan H.B., Moore E., Ramage D., Hampson S., y Arcas B.A., "Communication-efficient learning of deep networks from decentralized data," Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, 54, 1273-1282.
5. Diro A.A., Chilamkurti N., "Distributed attack detection scheme using deep learning approach for Internet of Things," Future Generation Computer Systems, 2018, 82, 761-768.
6. Preuveneers D., Rimmer V., Tsingenopoulos I., Spooren J., Joosen W., Ilie-Zudor E., "Chained anomaly detection models for federated learning: An intrusion detection case study," Applied Sciences, 2018, 8 (12), 2663.
7. Mothukuri V., Parizi R.M., Pouriyeh S., Huang Y., Dehghantanha A., Srivastava G., "A survey on security and privacy of federated learning," Future Generation Computer Systems, 2021, 115, 619-640.
8. Rahman S.A., Tout H., Talhi C., Mourad A., "Internet of Things intrusion detection: Centralized, on-device, or federated learning?," IEEE Network, 2020, 34 (6), 310-317.
9. Bhagoji A.N., Chakraborty S., Mittal P., Calo S., "Analyzing federated learning through an adversarial lens," Proceedings of the 36th International Conference on Machine Learning (ICML), 2019, 97, 634-643.
10. Blanchard P., El Mhamdi E.M., Guerraoui R., Stainer J., "Machine learning with adversaries: Byzantine tolerant gradient descent," Advances in Neural Information Processing Systems, 2017, 30, 119-129.
11. Meidan Y., Bohadana M., Mathov Y., Mirsky Y., Shabtai A., Breitenbacher D., Elovici Y., "N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders," IEEE Pervasive Computing, 2018, 17 (3), 12-22.
12. Moustafa N., Turnbull B., Choo K.K.R., "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things," IEEE Internet of Things Journal, 2019, 6 (3), 4815-4830.