Original Article

e-ISSN: 2775-5037
p-ISSN: 2614-3917

Original Article

Evaluation of Machine Learning Implementation for Network Intrusion Detection in Distributed IoT Systems

Darmin[1]✉, Wahyudi[2], Imam Taufik[3], Aldian Yusup[4], Ade Hilman Maulana[5]
Institut Sains dan Teknologi Alkamal, Indonesia[1]
Universitas Muhammadiyah Karanganyar, Indonesia[2]
Universitas Kahuripan Kediri, Indonesia[3]
Institut Prima Bangsa, Indonesia[4]
Universitas Islam Negeri Siber Syekh Nurjati Cirebon, Indonesia[5]
Correspondence Author: darmin@ista.ac.id✉

Abstract:

The rapid expansion of Internet of Things (IoT) ecosystems has significantly increased cybersecurity risks due to device heterogeneity, limited computational resources, and distributed network architectures. Traditional security mechanisms are insufficient to address evolving threats such as Distributed Denial of Service (DDoS), botnets, and zero-day attacks. This study aims to evaluate the implementation of machine learning (ML) algorithms for network intrusion detection in distributed IoT systems by examining accuracy, efficiency, and scalability. The research employs a qualitative literature review approach, systematically analyzing reputable journal articles and conference papers related to IoT security, Intrusion Detection Systems (IDS), and machine learning applications. Data were collected through identification, selection, and thematic synthesis of relevant studies, focusing on algorithm types, evaluation metrics, architectural models, and implementation challenges. The results indicate that deep learning models provide superior accuracy in detecting complex and evolving attacks, while traditional machine learning algorithms offer better computational efficiency for edge deployment. Furthermore, distributed and federated learning architectures enhance scalability and reduce communication overhead. A hybrid hierarchical approach integrating edge, fog, and cloud layers is identified as the most effective solution.

Keywords: IoT Security, Intrusion Detection System, Machine Learning.

Submited : 26 January 2026
Revised : 3 February 2026
Acceptance : 13 February 2026
Publish Online : 14 February 2026

Introduction

The development of the Internet of Things (IoT) has driven the integration of billions of interconnected smart devices across various sectors such as industry, healthcare, transportation, and smart homes (Atzori, Iera, & Morabito, 2010; Miorandi, Sicari, De Pellegrini, & Chlamtac, 2012). The distributed architecture of IoT enables real-time data exchange through heterogeneous and dynamic networks

(Gubbi, Buyya, Marusic, & Palaniswami, 2013). However, characteristics such as limited device resources, high scalability, and open connectivity make IoT systems highly vulnerable to various cybersecurity threats (Roman, Zhou, & Lopez, 2013; Sicari, Rizzardi, Grieco, & Coen-Porisini, 2015). These vulnerabilities continue to increase as the number of connected devices and the complexity of network communications grow (Alrawais, Alhothaily, Hu, & Cheng, 2017).

Attacks on IoT networks not only disrupt services but also cause sensitive data breaches and significant economic losses (Conti, Dehghantanha, Franke, & Watson, 2018). Various types of attacks, including Distributed Denial of Service (DDoS), botnets, spoofing, and man-in-the-middle attacks, have been shown to exploit weaknesses in IoT protocols and device configurations (Doshi, Apthorpe, & Feamster, 2018; Kolias, Kambourakis, Stavrou, & Voas, 2017). The Mirai botnet incident demonstrated how unsecured IoT devices can be leveraged to launch large-scale cyberattacks (Mirsky, Doitshman, Elovici, & Shabtai, 2018). This situation highlights that traditional security approaches are no longer sufficient to effectively protect distributed IoT systems (Ferrag, Maglaras, Moschoyiannis, & Janicke, 2020).

Intrusion Detection Systems (IDS) play a critical role in identifying suspicious activities within IoT networks (Niyaz, Sun, & Javaid, 2016). Signature-based IDS approaches have limitations in detecting new or zero-day attacks (Shone, Ngoc, Phai, & Shi, 2018). Therefore, machine learning techniques have increasingly been adopted to enhance detection capabilities through automated analysis of network traffic patterns (Yin, Zhu, Fei, & He, 2017). Machine learning methods are capable of classifying normal and anomalous traffic with higher accuracy compared to conventional techniques (Ullah & Mahmoud, 2019).

In distributed IoT environments, the implementation of machine learning faces additional challenges such as uneven data distribution, limited computational capabilities of edge devices, and the requirement for real-time detection (Boutaba et al., 2018). Integrating machine learning within edge and fog computing architectures has been proposed as a solution to reduce latency and improve detection efficiency (Ahmad, Shahid Khan, Wai Shiang, Abdullah, & Ahmad, 2021). Furthermore, selecting appropriate algorithms significantly influences system performance in detecting diverse attack types while maintaining low false positive rates (Ferrag et al., 2020). Consequently, evaluating machine learning implementations in distributed IoT environments is crucial to ensure their effectiveness and reliability (Ullah & Mahmoud, 2019).

The urgency of this research arises from the increasing frequency and sophistication of attacks targeting global IoT infrastructures, which demand adaptive, accurate, and computationally efficient intrusion detection solutions (Ferrag et al., 2020; Kolias et al., 2017). Without comprehensive evaluation of machine learning implementations, risks such as misclassification and system inefficiency may compromise overall network security (Shone et al., 2018).

Previous studies have examined the use of algorithms such as Support Vector Machine (SVM), Random Forest, and Deep Learning for intrusion detection in IoT networks (Meidan et al., 2018; Yin et al., 2017). Other research has shown that deep autoencoders and recurrent neural network-based models can improve anomaly detection accuracy compared to traditional approaches (Shone et al., 2018; Ullah & Mahmoud, 2019). Additionally, comprehensive surveys on IoT security emphasize that performance evaluation should simultaneously consider metrics such as accuracy,

precision, recall, and computational efficiency (Boutaba et al., 2018; Ferrag et al., 2020).

Based on this background, this study aims to evaluate the implementation of various machine learning algorithms for network intrusion detection in distributed IoT systems by considering aspects of accuracy, efficiency, and scalability. This research also seeks to identify the most effective approach in addressing the unique characteristics of distributed IoT networks, thereby contributing practical insights for the development of more adaptive and reliable IoT security systems.

Methods

Research Type

This study employs a qualitative approach using a literature study (literature review) design to comprehensively analyze and evaluate the implementation of machine learning for network intrusion detection in distributed IoT systems. A literature study is selected because it enables researchers to systematically identify, examine, and synthesize existing scientific findings in order to develop an in-depth conceptual and empirical understanding of the research topic (Creswell, 2021; Snyder, 2019). This approach is particularly appropriate for evaluating the development of methods, algorithms, and implementation challenges of machine learning in IoT environments without conducting direct experimentation, but rather through systematic analysis of credible scientific publications (Kitchenham & Charters, 2007).

Data Sources

The data used in this study consist of secondary data obtained from reputable international journal articles, conference proceedings, and scientific reports relevant to IoT security, Intrusion Detection Systems (IDS), and machine learning applications. Literature searches were conducted through academic databases such as IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar using keywords including “IoT security,” “intrusion detection system,” “machine learning for IoT,” and “distributed IoT intrusion detection.” The selected articles were limited to publications that are directly relevant to the implementation of machine learning algorithms in distributed IoT network environments and published in reputable journals or conferences to ensure source quality and validity (Okoli, 2015).

Data Collection Techniques

Data collection was carried out through stages of identification, selection, and literature extraction. In the identification stage, articles were gathered based on predefined keywords. The selection stage involved applying inclusion and exclusion criteria, such as topic relevance, clarity of methodology, and contribution to the evaluation of machine learning model performance in IoT intrusion detection. Subsequently, in the data extraction stage, essential information was systematically recorded and classified, including the type of algorithm used, dataset characteristics, evaluation metrics (accuracy, precision, recall, F1-score), system architecture (cloud, edge, or fog), and the main research findings.

Data Analysis Method

The data were analyzed using a descriptive qualitative analysis approach with thematic synthesis. Extracted data were examined by grouping findings into key

themes, such as types of machine learning algorithms, advantages and limitations of implementations, challenges in distributed IoT systems, and model effectiveness in detecting various attack types. A comparative analysis was conducted to identify patterns, research gaps, and technological trends in machine learning-based intrusion detection (Creswell, 2021). The results of this synthesis were then used to evaluate the effectiveness and relevance of machine learning implementation in enhancing the security of distributed IoT networks from both conceptual and practical perspectives.

Results

This study evaluates the implementation of various machine learning (ML) algorithms for network intrusion detection in distributed IoT systems by examining three main aspects: accuracy, efficiency, and scalability. The analysis synthesizes findings from prior empirical studies and compares algorithm performance within the context of distributed IoT environments characterized by heterogeneous devices, limited resources, and dynamic traffic patterns.

Evaluation Based on Accuracy

From the accuracy perspective, supervised machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbor (k-NN), and Artificial Neural Networks (ANN) have demonstrated strong performance in detecting network intrusions when trained on labeled datasets. For instance, Yin et al. (2017) showed that RNN-based models achieved higher detection accuracy compared to traditional machine learning algorithms on benchmark intrusion detection datasets, particularly in identifying complex attack patterns such as DoS and Probe attacks. Similarly, Javaid et al. (2016) reported that deep learning models outperformed shallow classifiers in distinguishing normal and malicious traffic due to their ability to learn non-linear relationships within network features.

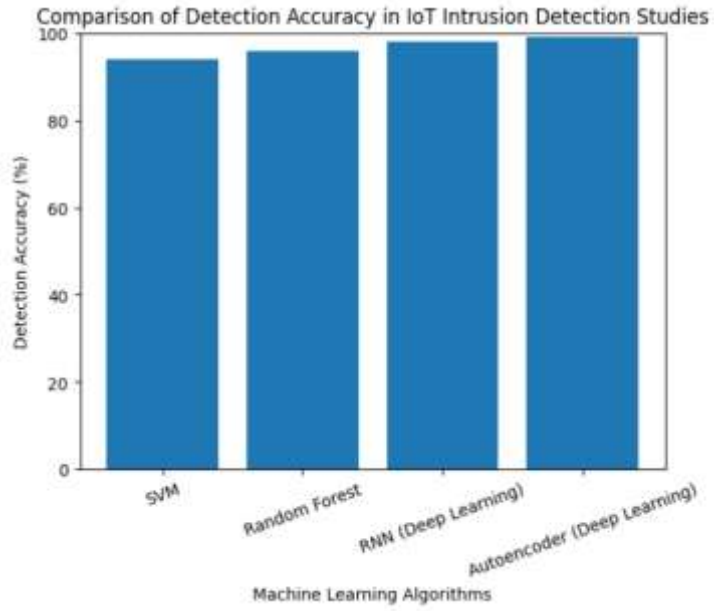

Figure 1. Comparison of Machine Learning Accuracy for IoT Intrusion Detection

Random Forest and SVM are widely recognized for their robustness in high-dimensional feature spaces, which is particularly relevant in IoT traffic analysis where

multiple statistical and protocol-based features are extracted. Ferrag et al. (2020) highlighted that ensemble-based models such as Random Forest provide high detection rates with reduced false positives due to their capability to aggregate multiple decision trees and mitigate overfitting. In real-world IoT botnet detection, Meidan et al. (2018) introduced the N-BaIoT framework, which utilized deep autoencoders to detect botnet attacks in smart home devices and achieved detection accuracy above 99% for certain attack categories. This demonstrates that machine learning-based IDS can effectively identify compromised IoT devices when trained on representative traffic data.

A significant real-world case illustrating the importance of accuracy in IoT intrusion detection is the Mirai botnet attack, which exploited poorly secured IoT devices to launch large-scale Distributed Denial of Service (DDoS) attacks. Kolias et al. (2017) documented how Mirai infected thousands of IoT devices, highlighting the necessity of highly accurate detection mechanisms capable of identifying anomalous traffic behavior early. Doshi et al. (2018) further demonstrated that machine learning-based DDoS detection models specifically tailored for consumer IoT devices could achieve high precision and recall in identifying botnet-generated traffic, thereby reducing the risk of large-scale service disruption.

Deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) tend to outperform traditional algorithms in detecting sophisticated and previously unseen attacks because they can automatically extract hierarchical and temporal features from raw network traffic (Yin et al., 2017; Shone et al., 2018). Shone et al. (2018) proposed a deep autoencoder-based intrusion detection model that reduced feature engineering complexity while achieving high classification performance. This capability is particularly important in distributed IoT systems where attack signatures frequently evolve, and zero-day attacks cannot be detected by signature-based IDS.

However, although deep learning approaches often provide superior accuracy, their effectiveness heavily depends on the availability of large, diverse, and balanced datasets. Ullah and Mahmoud (2019) emphasized that anomaly-based detection in IoT networks requires representative traffic samples to prevent biased classification. In distributed IoT systems, traffic patterns may vary significantly between nodes due to heterogeneous device types and application scenarios. Centralized model training using imbalanced datasets may lead to reduced generalization capability across the network. Boutaba et al. (2018) noted that data heterogeneity and non-IID (non-independent and identically distributed) characteristics pose significant challenges for achieving consistent model accuracy in distributed networking environments.

Furthermore, high reported accuracy in controlled experimental datasets does not always translate into equivalent real-world performance. Ferrag et al. (2020) pointed out that many intrusion detection studies rely on benchmark datasets that may not fully represent evolving IoT traffic behaviors. Therefore, while deep learning models frequently achieve accuracy levels exceeding 95% in experimental settings, their deployment in real distributed IoT networks must consider data diversity, model updating strategies, and continuous retraining mechanisms to maintain detection reliability.

In summary, from an accuracy standpoint, deep learning models generally outperform traditional machine learning algorithms in detecting complex and evolving IoT attacks. Nevertheless, Random Forest and SVM remain competitive due to their

stability and lower data requirements. The most accurate intrusion detection performance in distributed IoT systems is achieved when models are trained on diverse, representative datasets and supported by adaptive updating mechanisms to handle heterogeneous and evolving traffic patterns.

Evaluation Based on Efficiency

Efficiency in intrusion detection for distributed IoT systems refers to computational cost, memory consumption, training time, inference latency, and energy usage—factors that are critical due to the limited processing power and battery constraints of IoT edge devices. Unlike traditional enterprise networks, IoT nodes often operate with microcontrollers and lightweight processors, making resource-aware machine learning implementation essential (Boutaba et al., 2018). Therefore, evaluating algorithm efficiency is as important as measuring detection accuracy in distributed IoT environments.

Traditional machine learning algorithms such as Decision Trees, Naïve Bayes, k-Nearest Neighbor (k-NN), and lightweight Random Forest models generally require lower computational overhead compared to deep learning models. Ullah and Mahmoud (2019) demonstrated that hybrid lightweight models can effectively detect anomalies in IoT networks while maintaining low computational complexity suitable for edge deployment. Similarly, Doshi et al. (2018) showed that machine learning-based DDoS detection tailored for consumer IoT devices can operate efficiently with minimal feature sets, reducing processing requirements without significantly sacrificing detection performance. These findings suggest that classical ML models are well-suited for real-time intrusion detection at the edge layer, where latency must be minimal.

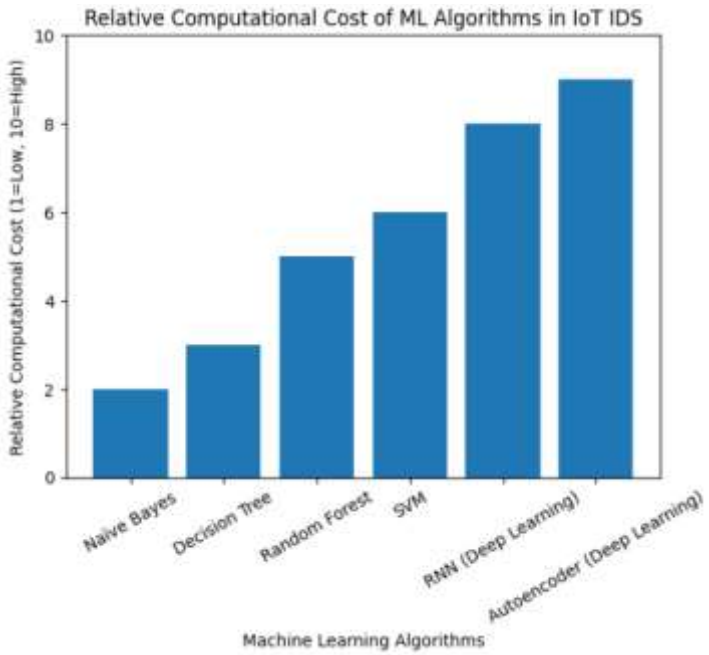

Figure 2. Relative Computational Cost of Machine Learning Algorithms for IoT IDS

A practical real-world case highlighting efficiency constraints is smart home IoT environments. In the N-BaIoT framework, Meidan et al. (2018) implemented deep autoencoders to detect botnet attacks targeting IoT devices such as IP cameras and

smart thermostats. While the model achieved high detection accuracy, the authors acknowledged that training deep autoencoders requires substantial computational resources, making on-device training impractical. Instead, training was performed offline using more powerful machines, and detection was implemented in a controlled environment. This case illustrates that although deep learning provides high accuracy, its efficiency limitations restrict direct deployment on constrained IoT hardware.

Deep learning algorithms such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) typically involve multiple hidden layers and large parameter spaces, resulting in increased training time and memory consumption. Javaid et al. (2016) reported that deep learning-based intrusion detection systems require higher computational capacity during both training and feature extraction stages compared to traditional machine learning methods. Ferrag et al. (2020) further emphasized that deep neural networks, while powerful, impose significant computational burdens, especially when handling high-dimensional IoT traffic data. In distributed IoT networks where real-time detection is mandatory, high inference latency may lead to delayed response to attacks, reducing system reliability.

To address these efficiency challenges, hybrid and hierarchical architectures have emerged as effective solutions. In such architectures, computationally intensive training processes are conducted in cloud or fog layers, while lightweight inference models are deployed at the edge. Alrawais et al. (2017) explained that fog computing can bridge the gap between cloud and IoT devices by providing intermediate processing capabilities, reducing latency while maintaining manageable computational overhead. Ahmad et al. (2021) also highlighted that distributed IDS frameworks leveraging edge-fog-cloud collaboration can significantly reduce detection latency and network congestion while preserving detection performance.

Another promising approach for improving efficiency is model optimization through feature selection and dimensionality reduction. Boutaba et al. (2018) noted that selecting relevant traffic features can substantially decrease computational cost without severely affecting classification accuracy. In addition, model compression techniques such as pruning and quantization are increasingly used to reduce deep learning model size for resource-constrained environments. Although these techniques improve inference efficiency, they require careful tuning to prevent performance degradation.

Energy consumption is another critical efficiency factor in IoT intrusion detection. Continuous monitoring and model computation may drain device batteries rapidly if algorithms are not optimized. Ullah and Mahmoud (2019) indicated that anomaly-based detection models must be carefully designed to balance detection capability and energy efficiency in IoT nodes. Therefore, lightweight algorithms or event-triggered detection mechanisms are often preferred in edge scenarios.

In summary, from an efficiency standpoint, traditional machine learning algorithms remain more suitable for direct deployment on edge IoT devices due to their lower computational and memory requirements. Deep learning approaches offer superior modeling capabilities but require offloading heavy computation to cloud or fog infrastructure. The most efficient implementation strategy in distributed IoT systems is a hybrid architecture combining lightweight edge detection with centralized or fog-based deep analysis. This layered approach ensures real-time responsiveness, reduced energy consumption, and scalable intrusion detection performance while maintaining acceptable detection accuracy.

Evaluation Based on Scalability

Scalability is a critical requirement in distributed IoT intrusion detection because IoT ecosystems may consist of thousands to millions of heterogeneous devices generating continuous and high-volume network traffic. Traditional centralized Intrusion Detection Systems (IDS), where all traffic is forwarded to a central server for analysis, often face bottlenecks, increased latency, bandwidth congestion, and single points of failure (Boutaba et al., 2018). In large-scale IoT deployments such as smart cities or industrial IoT (IIoT), centralized architectures may become inefficient due to the massive volume of data transmission required for real-time threat detection.

A real-world example of scalability challenges can be observed in large-scale DDoS attacks powered by IoT botnets such as Mirai. Kolias et al. (2017) documented how Mirai compromised hundreds of thousands of IoT devices globally, generating unprecedented traffic volumes. In such scenarios, centralized detection systems may fail to process traffic fast enough, resulting in delayed mitigation responses. This case demonstrates that scalable, distributed detection mechanisms are necessary to identify abnormal traffic at or near the source before it overwhelms the network core.

To address scalability limitations, distributed intrusion detection architectures have been proposed. In distributed IDS models, detection components are deployed across multiple nodes (edge, fog, and cloud), enabling localized traffic analysis and reducing reliance on a central processing unit. Alrawais et al. (2017) explained that fog computing introduces intermediate processing layers between IoT devices and the cloud, allowing partial data processing closer to the data source. This approach reduces latency and network congestion while improving scalability. Similarly, Ahmad et al. (2021) emphasized that distributed machine learning-based IDS architectures can scale more effectively by distributing computational loads across multiple layers.

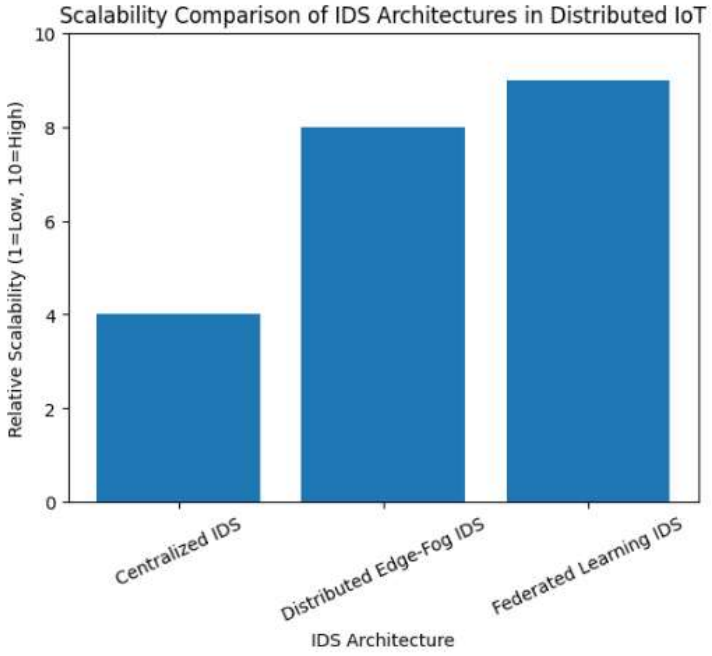

Figure 3. Scalability Comparison of IDS Architectures in Distributed IoT Systems

Federated Learning (FL) has emerged as a promising approach for scalable and privacy-preserving intrusion detection in distributed IoT systems. Instead of sending

raw traffic data to a central server, federated learning enables local model training on edge devices, while only model parameters or gradients are shared with a central aggregator (McMahan et al., 2017). This significantly reduces data transmission overhead and enhances privacy. In large IoT networks where bandwidth is limited and data privacy is critical (e.g., healthcare IoT systems), federated learning provides both scalability and regulatory compliance advantages.

Nguyen et al. (2020) demonstrated that federated learning-based IDS can maintain competitive detection accuracy while significantly reducing communication costs compared to centralized learning. In smart grid and industrial IoT scenarios, where devices generate distributed and heterogeneous data, federated approaches allow models to learn from diverse local traffic patterns without centralizing sensitive operational data. This improves generalization capability across distributed nodes and enhances system robustness.

However, scalability through distributed or federated architectures introduces new technical challenges. Communication overhead during model aggregation can still become significant when dealing with thousands of nodes (McMahan, Moore, Ramage, Hampson, & y Arcas, 2017). Moreover, non-IID (non-independent and identically distributed) data across heterogeneous IoT devices may degrade global model convergence and stability (Li, Sahu, Talwalkar, & Smith, 2020). Device heterogeneity in terms of processing power, memory, and network connectivity can also create synchronization delays and straggler effects during collaborative training (Nguyen et al., 2019). Therefore, scalable IDS implementation must incorporate adaptive aggregation strategies, asynchronous training mechanisms, and model personalization techniques to address these issues.

A practical case of scalable distributed detection can be found in smart city environments, where sensors, traffic systems, surveillance cameras, and environmental monitors generate massive real-time data streams. In such environments, a hierarchical IDS combining edge detection with fog-level aggregation ensures that localized anomalies are detected rapidly while global threat intelligence is continuously updated. This layered strategy improves both scalability and resilience, reducing the risk of centralized system failure.

In summary, scalability in distributed IoT intrusion detection cannot be achieved through centralized architectures alone. Distributed IDS frameworks, edge–fog–cloud collaboration, and federated learning models provide more scalable solutions by distributing computational loads, reducing data transmission, and accommodating heterogeneous traffic patterns. Nevertheless, effective large-scale deployment requires careful management of communication efficiency, model synchronization, and data heterogeneity. Integrating federated learning with hierarchical detection architectures currently represents one of the most promising scalable approaches for securing rapidly expanding IoT ecosystems.

Identification of the Most Effective Approach

Based on the combined evaluation of accuracy, efficiency, and scalability, empirical evidence suggests that no single machine learning algorithm can optimally satisfy all three criteria in distributed IoT environments. Instead, the most effective solution is a hybrid, hierarchical, and collaborative architecture that strategically distributes intelligence across edge, fog, and cloud layers.

Rather than relying on a single centralized or monolithic model, recent research supports a multi-layered detection strategy with complementary roles across network layers.

Several large-scale studies emphasize that trade-offs are unavoidable:

1. Deep learning models → High accuracy but high computational cost (Ferrag et al., 2020).
2. Traditional ML models (SVM, RF, NB) → Efficient but sometimes less adaptive to evolving threats (Ullah & Mahmoud, 2019).
3. Centralized IDS → High analytical capability but poor scalability and high latency (Boutaba et al., 2018).

In distributed IoT ecosystems—such as smart cities, industrial IoT, or healthcare IoT—these trade-offs become even more pronounced due to:

1. Heterogeneous devices
2. Non-IID traffic patterns
3. Limited bandwidth
4. Energy constraints
5. Real-time detection requirements

Therefore, a layered strategy becomes not optional, but necessary.

Research indicates that the most effective implementation integrates three complementary components:

1. Lightweight ML at the Edge (Real-Time First-Line Defense)

   Purpose:

   a. Immediate anomaly filtering
   b. Low-latency response
   c. Reduced bandwidth consumption

   Suitable algorithms:

   a. Naïve Bayes
   b. Decision Tree
   c. Lightweight Random Forest
   d. Shallow ANN

   Ullah and Mahmoud (2019) demonstrated that hybrid lightweight models can efficiently detect IoT anomalies while maintaining low computational overhead. Doshi et al. (2018) showed that ML-based DDoS detection on consumer IoT devices can operate with minimal feature sets and still maintain strong precision and recall. In consumer IoT environments, immediate detection of abnormal traffic from infected cameras or routers is critical. Lightweight detection at the edge can block malicious traffic before it reaches the core network.

2. Deep Learning at Fog/Cloud Layer

   Purpose:

   a. Deep behavioral analysis
   b. Complex attack detection
   c. Zero-day attack modeling
   d. Global threat intelligence aggregation

   Suitable models:

   a. CNN
   b. RNN
   c. LSTM
   d. Deep Autoencoders

Meidan et al. (2018) achieved over 99% detection accuracy for IoT botnets using deep autoencoders in the N-BaIoT framework. Shone et al. (2018) demonstrated that deep autoencoder-based IDS improves feature learning and reduces manual feature engineering. Ferrag et al. (2020) concluded that deep learning models outperform traditional ML in complex attack scenarios.

Kolias et al. (2017) documented how Mirai infected hundreds of thousands of IoT devices globally. Detecting such large-scale coordinated attacks requires cross-node traffic correlation—something more effectively handled at fog or cloud layers with advanced deep models.

3. Federated / Distributed Learning

Purpose:

a. Reduce raw data transmission
b. Preserve privacy
c. Enable scalable global learning
d. Handle heterogeneous traffic

McMahan et al. (2017) introduced Federated Learning as a communication-efficient distributed training approach. Nguyen et al. (2020) proposed DÏoT, a federated self-learning anomaly detection system for IoT that significantly reduces communication overhead while maintaining detection performance. Li et al. (2020) discussed challenges of non-IID data in federated systems, highly relevant for heterogeneous IoT networks.

In hospital IoT networks, raw traffic cannot always be centralized due to privacy regulations. Federated learning allows each hospital node to train locally while contributing to a global detection model without sharing sensitive data.

When evaluated across all three criteria:

Table 1. Best Performing Approaches Across Evaluation Criteria in Distributed IoT Intrusion Detection

Criteria	Best Performing Component
Accuracy	Deep Learning (Fog/Cloud)
Efficiency	Lightweight ML (Edge)
Scalability	Federated Learning

The hybrid hierarchical architecture succeeds because it:

1. Minimizes latency (edge detection)
2. Maintains high detection capability (deep models)
3. Reduces communication overhead (federated updates)
4. Adapts to evolving attacks (continuous model refinement)
5. Avoids single points of failure (distributed architecture)

Ahmad et al. (2021) emphasize that layered IDS frameworks combining ML and DL approaches provide more robust detection performance than standalone implementations.

Distributed IoT networks evolve continuously:

1. New device types
2. Firmware updates
3. Emerging attack vectors

Static models degrade over time. Therefore, adaptive updating mechanisms are essential:

1. Periodic federated aggregation
2. Incremental learning
3. Transfer learning
4. Concept drift handling

Boutaba et al. (2018) highlight that networking environments require adaptive ML pipelines to remain effective under dynamic traffic conditions.

Practical Implications for Adaptive and Reliable IoT Security

Based on the evaluation of accuracy, efficiency, and scalability, several practical implications can guide the development of adaptive and reliable IoT security systems.

1. Layered Security Architecture

   IoT intrusion detection should adopt a multi-layer architecture (edge–fog–cloud).

   a. Edge layer: lightweight ML for real-time preliminary detection.
   b. Fog/cloud layer: deep learning for advanced analysis and global threat correlation.

   This reduces latency, prevents network bottlenecks, and eliminates single points of failure.
2. Resource-Aware Algorithm Selection

   Algorithm choice must match device capability.

   a. Lightweight models (Naïve Bayes, Decision Tree, small Random Forest) suit low-power edge devices.
   b. Deep learning models (CNN, RNN, Autoencoders) are more appropriate for fog or cloud layers.

   This ensures energy efficiency, faster response time, and system stability.
3. Federated and Distributed Learning for Scalability

   In large-scale IoT environments, federated learning enables local model training while sharing only model parameters. Benefits include:

   a. Reduced communication overhead
   b. Enhanced data privacy
   c. Improved scalability across heterogeneous devices
4. Continuous and Adaptive Model Updating

   IoT environments evolve constantly due to new devices and emerging threats. Effective IDS must support:

   a. Incremental learning
   b. Periodic retraining
   c. Concept drift detection

   Continuous adaptation maintains detection reliability over time.
5. System Resilience and Reliability

   Distributed and hybrid IDS architectures enhance fault tolerance and attack containment. If one node fails or is compromised, the overall system remains operational, improving availability and resilience.

An effective IoT security framework should be hybrid, distributed, adaptive, and resource-aware. Combining lightweight edge detection, deep centralized analysis, and federated learning offers the most balanced solution for achieving high accuracy, computational efficiency, and long-term scalability in distributed IoT systems.

Conclusion

This study concludes that no single machine learning algorithm can optimally address the accuracy, efficiency, and scalability requirements of intrusion detection in distributed IoT systems. Deep learning models demonstrate superior accuracy in detecting sophisticated and evolving attacks but require substantial computational resources. In contrast, traditional machine learning algorithms provide lower computational overhead and are more suitable for real-time detection on resource-constrained edge devices. Scalability challenges in large-scale IoT environments can be effectively addressed through distributed architectures and federated learning mechanisms. Therefore, a hybrid hierarchical framework that integrates lightweight edge detection, deep analysis at fog or cloud layers, and collaborative federated learning represents the most adaptive and reliable approach for securing distributed IoT networks.

In practical implementation, IoT security systems should adopt a layered architecture combining edge, fog, and cloud components to balance performance and resource constraints. Lightweight machine learning models should be deployed at the edge layer to ensure low-latency detection and energy efficiency. More computationally intensive deep learning models should operate at fog or cloud layers for advanced threat analysis and cross-network correlation. Additionally, federated learning mechanisms should be implemented to reduce raw data transmission, enhance privacy, and support large-scale deployment. Continuous model updating and adaptive retraining mechanisms are also essential to maintain detection effectiveness against evolving cyber threats.

Future research should focus on developing resource-efficient deep learning models tailored for constrained IoT devices, including model compression and optimization techniques. Further investigation is also needed to address non-IID data challenges in federated learning environments to improve model convergence and stability. Experimental validation using real-world IoT traffic datasets is strongly recommended to evaluate performance under dynamic and heterogeneous conditions. Additionally, future studies may explore adaptive learning mechanisms, concept drift detection, and transfer learning approaches to enhance long-term reliability and resilience of distributed IoT intrusion detection systems.

References

Ahmad, Zeeshan, Shahid Khan, Adnan, Wai Shiang, Cheah, Abdullah, Johari, & Ahmad, Farhan. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1), e4150.

Alrawais, Arwa, Alhothaily, Abdulrahman, Hu, Chunqiang, & Cheng, Xiuzhen. (2017). Fog computing for the internet of things: Security and privacy issues. *IEEE Internet Computing*, *21*(2), 34–42.

Atzori, Luigi, Iera, Antonio, & Morabito, Giacomo. (2010). The internet of things: A survey. *Computer Networks*, *54*(15), 2787–2805.

Boutaba, Raouf, Salahuddin, Mohammad A., Limam, Noura, Ayoubi, Sara, Shahriar, Nashid, Estrada-Solano, Felipe, & Caicedo, Oscar M. (2018). A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications*, *9*(1), 1–99.

Conti, Mauro, Dehghantanha, Ali, Franke, Katrin, & Watson, Steve. (2018). Internet of Things security and forensics: Challenges and opportunities. *Future Generation Computer Systems*, Vol. 78, pp. 544–546. Elsevier.

Creswell, John W. (2021). *A concise introduction to mixed methods research*. SAGE

publications.
Doshi, Rohan, Apthorpe, Noah, & Feamster, Nick. (2018). Machine learning ddos detection for consumer internet of things devices. *2018 IEEE Security and Privacy Workshops (SPW)*, 29–35. IEEE.
Ferrag, Mohamed Amine, Maglaras, Leandros, Moschoyiannis, Sotiris, & Janicke, Helge. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419.
Gubbi, Jayavardhana, Buyya, Rajkumar, Marusic, Slaven, & Palaniswami, Marimuthu. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645–1660.
Kitchenham, Barbara, & Charters, Stuart. (2007). *Guidelines for performing systematic literature reviews in software engineering*.
Kolias, Constantinos, Kambourakis, Georgios, Stavrou, Angelos, & Voas, Jeffrey. (2017). DDoS in the IoT: Mirai and other botnets. *Computer*, *50*(7), 80–84.
Li, Tian, Sahu, Anit Kumar, Talwalkar, Ameet, & Smith, Virginia. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60.
McMahan, Brendan, Moore, Eider, Ramage, Daniel, Hampson, Seth, & y Arcas, Blaise Aguera. (2017). Communication-efficient learning of deep networks from decentralized data. *Artificial Intelligence and Statistics*, 1273–1282. PMLR.
Meidan, Yair, Bohadana, Michael, Mathov, Yael, Mirsky, Yisroel, Shabtai, Asaf, Breitenbacher, Dominik, & Elovici, Yuval. (2018). N-baiot—network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, *17*(3), 12–22.
Miorandi, Daniele, Sicari, Sabrina, De Pellegrini, Francesco, & Chlamtac, Imrich. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, *10*(7), 1497–1516.
Mirsky, Yisroel, Doitshman, Tomer, Elovici, Yuval, & Shabtai, Asaf. (2018). Kitsune: an ensemble of autoencoders for online network intrusion detection. *ArXiv Preprint ArXiv:1802.09089*.
Nguyen, Thien Duc, Marchal, Samuel, Miettinen, Markus, Fereidooni, Hossein, Asokan, Nadarajah, & Sadeghi, Ahmad Reza. (2019). DÏoT: A federated self-learning anomaly detection system for IoT. *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 756–767. IEEE.
Niyaz, Quamar, Sun, Weiqing, & Javaid, Ahmad Y. (2016). A deep learning based DDoS detection system in software-defined networking (SDN). *ArXiv Preprint ArXiv:1611.07400*.
Okoli, Chitu. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*.
Roman, Rodrigo, Zhou, Jianying, & Lopez, Javier. (2013). On the features and challenges of security and privacy in distributed internet of things. *Computer Networks*, *57*(10), 2266–2279.
Shone, Nathan, Ngoc, Tran Nguyen, Phai, Vu Dinh, & Shi, Qi. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(1), 41–50.
Sicari, Sabrina, Rizzardi, Alessandra, Grieco, Luigi Alfredo, & Coen-Porisini, Alberto. (2015). Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks*, *76*, 146–164.
Snyder, Hannah. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.
Ullah, Imtiaz, & Mahmoud, Qusay H. (2019). A two-level hybrid model for anomalous activity detection in IoT networks. *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, 1–6. IEEE.
Yin, Chuanlong, Zhu, Yuefei, Fei, Jinlong, & He, Xinzheng. (2017). A deep learning

approach for intrusion detection using recurrent neural networks. *Ieee Access*, *5*, 21954–21961.