ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

Scalable Privacy-Preserving Cyber Defense: Federated Self-Supervised Learning for Zero-Day Threat Detection in Critical Infrastructure

Sadiya Afrin[1]; Jawad Sarwar[2]

[1]School of IT Washington University of Science and Technology
[2]School of IT Washington University of Science and Technology

Publication Date: 2026/05/07

Abstract: The high rate of digitalization of major infrastructural systems, such as energy grids, transportation systems, healthcare services, and industrial control systems, has greatly exposed them to advanced cyberattacks. Of these threats, zero-day attacks are especially dangerous because they have an unknown signature and cannot be detected by traditional signature-based detection mechanisms before they cause any damage. Simultaneously, centralized security analytics solutions also come with significant privacy, regulatory, and operational risks, since sensitive operational data cannot be easily distributed across distributed facilities. The proposed study suggests a Federated Self-Supervised Federated Defense framework that combines both federated learning and self-supervised repair learning to allow privacy-preserving collaborative zero-day threat detection across geographically distributed network nodes of infrastructure. This is made possible by the suggested architecture, which enables local systems to utilize models trained on-site, while only sharing some encrypted model parameters, ensuring confidentiality is maintained and regulatory compliance is upheld. Self-supervised pretraining is more sensitive to anomalies by learning inherent behavioral patterns based on unlabeled network and system telemetry, which is more useful for generalization to unseen attacks. Experimental analysis with simulated infrastructure datasets that are distributed shows that the detection accuracy is higher, the number of false positives is lower, and the communication overhead is less than with centralized and purely supervised baselines. The framework is also resistant to data heterogeneity and adversarial manipulation via secure aggregation and adaptive model updates. Findings indicate that federated self-supervised learning can be utilized to substantially enhance collective cyber defense without compromising privacy or operational independence. This study highlights a scalable and reliable future for next-generation smart surveillance of distributed critical infrastructure sites.

Keywords: *Federated Learning, Self-Supervised Learning, Zero-Day Attack Detection, Privacy-Preserving Cybersecurity, Critical Infrastructure Protection.*

How to Cite: Sadiya Afrin; Jawad Sarwar (2026) Scalable Privacy-Preserving Cyber Defense: Federated Self-Supervised Learning for Zero-Day Threat Detection in Critical Infrastructure. *International Journal of Innovative Science and Research Technology*, 11(5), 10-19. https://doi.org/10.38124/ijisrt/26may248

I. INTRODUCTION

➢ *Background and Motivation*

The energy grid, transportation systems, healthcare platforms, and industrial control environments are examples of critical infrastructure systems that are the backbone of contemporary society. The growing dependence on interdependent digital technologies, cloud services, and intelligent automation has immensely improved operational efficiency, scalability, and real-time decision-making. But the same digital transformation has also increased the cyber attack surface so that critical infrastructure systems are now extremely appealing targets for advanced cyber adversaries. Any damage to these systems can lead to substantial economic losses, a threat to human safety, and compromised national security, making cybersecurity both an operational and strategic issue [1], [5], [15].

Conventional approaches to cybersecurity defenses implemented in critical infrastructure settings have been mostly based on perimeter-based security models and signature-based intrusion detection systems. Although useful when faced with well-known threats, they cannot stand up to fast-developing methods of attack and unfamiliar vulnerabilities. Zero-day attacks that take advantage of formerly unknown vulnerabilities in software or systems are especially tough since they get through signature-based protection and are usually not detected until large-scale losses are realized [3], [8]. The increasing diversity and intricacy of critical infrastructure networks make it increasingly difficult to detect them in a timely manner, with attack signatures

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

potentially being different at the spread of distributed sites of operations [16].

Simultaneously, centralized security analytics solutions, where the information contributed by several nodes of an infrastructure is combined into one analysis service, are characterized by a grave set of privacy, regulatory, and operational challenges. Critical infrastructure systems can generate sensitive operational data that may include proprietary information, safety-critical parameters, or personally identifiable data that are not permitted to be freely shared because of regulatory limitations and national security concerns [7], [15]. These restrictions add a natural contradiction between the collaborative threat intelligence requirements and the necessity of upholding data sovereignty and confidentiality in the distributed infrastructure setting.

➢ *Zero-Day Detection and Privacy Preservation Problems:*

There is a distinct set of technical and organizational issues in identifying zero-day cyber threats in critical infrastructure systems. Contrary to traditional cyber environments, the infrastructure systems produce highly specialized and domain-specific telemetry data, which can be defined by very low false-positive tolerance and very demanding real-time operational constraints [12], [16]. Intrusion detection methods using machine learning and deep learning technologies have been demonstrated to perform well in detecting abnormal traffic, but most of the models require labeled datasets, which are unavailable or sparse in a zero-day setting [8], [11], [18]. This can lead to poor generalization of supervised learning techniques to the unknown pattern of attack.

Unsupervised and anomaly-based detection techniques overcome this shortcoming by learning the normal behavior of the system and marking an anomaly as a possible threat. Although these techniques enhance sensitivity to unidentified attacks, they are often characterized by high false alarm rates and weak scalability to use in heterogeneous contexts [3], [19]. The varying network structure, load distribution, and the performance of different devices across locations of the infrastructure can severely impair the performance of detection when models that are trained in a single setting are transferred to different locations.

There is a further complexification brought about by privacy preservation. Training models centrally needs raw data aggregation, which, in turn, subjects the operators of critical infrastructure to regulatory risks, data security risks, and insider risks [4], [7]. Partially addressing these risks are secure data-sharing systems like blockchain-based solutions and differential privacy methods, which, however, in general and particularly when used alone, require substantial computational costs or do not improve model utility [4], [7]. As a result, cybersecurity frameworks that would facilitate collaborative learning among distributed infrastructure systems without data exchange are urgently needed.

Federated learning is a new potential paradigm that can be used to resolve this problem by providing decentralized model training and sharing only model parameters, as opposed to raw data [9], [13], [17]. Nevertheless, even modern federated learning methods are based on the assumption that there is a large amount of labeled data and do not always work with non-identically distributed data that frequently occur in critical infrastructure settings [17], [23]. These shortcomings underscore the necessity to have more adaptive learning methods that would be able to derive meaningful representations on unlabeled data and still provide privacy assurances.

➢ *Research Objectives and Contributions:*

To address the limitations of current cybersecurity methods, this study offers a Federated Self-Supervised Cyber Defense framework that can support privacy-constrained zero-day threat detection in a distributed system of critical infrastructure. The fundamental purpose behind the given framework is to combine federated learning with self-supervised representation learning to collectively learn strong anomaly detection models without having to share sensitive operational information or make use of labelled attack images.

The methods of self-supervised learning have also been shown to be highly effective in learning quality feature representations with unlabeled data using intrinsic-based data structure and pretext problems [14], [20], [22], [24]. Self-supervised models can be trained locally on infrastructure nodes when combined with federated learning to obtain site-specific behavioral patterns, and something that combines with a globally optimized detection model to provide secure parameter aggregation [10], [21]. This combined methodology provides generalization to invisible threats and reduces the effects of heterogeneity of data within distributed settings.

This study made three major contributions. First, it presents a new federated self-supervised learning design targeting critical infrastructure cybersecurity, which deals with the issue of zero-day detection and a tight data privacy budget. Second, it proves that self-supervised representation learning enhances the sensitivity of anomalies and robustness in federated environments without affecting false positives. Third, it offers a comprehensive experimental assessment of the offered framework, as compared to centralized, supervised, and unsupervised baselines, which would reveal the enhanced detection accuracy, communication efficiency, and resilience to adversarial conditions.

This work contributes to the state of the art in privacy-preserving cyber defense by bridging federated learning, self-supervised learning, and critical infrastructure protection, as well as providing a scalable line of work in moving towards intelligent and collaborative security monitoring in the next-generation infrastructure systems.

II. LITERATURE REVIEW

➢ *Conventional and Zero-Day Intrusion Detection Methodologies:*

Early cyberspace defense against enterprises and industrial environments was mainly constructed on signature-

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

based intrusion discovery systems and rule-based firewalls, which were based on established attack patterns. Although these mechanisms can still be effective in the detection of known exploits, they are naturally reactive and not very effective against new or obfuscated threats. In critical infrastructure systems, where availability and safety are the most important, these reactive defenses cannot be considered sufficient, as attackers often use the vulnerabilities that were not known previously to evade detection. In turn, research has become more focused on anomaly-based and learning-based approaches that can be used to identify deviations of normal behaviour as opposed to going through fixed signatures [3], [8].

Unsupervised detection techniques are one of the most ancient efforts made to solve the zero-day threat by modeling baseline traffic or system behavior. Anomaly detection at the port, such as anomaly detection of deviation in communication patterns, can also detect suspicious activity without using labeled attack information [3]. These solutions are more sensitive to unknown attacks, but they tend to produce high levels of false positives because network environments are dynamic and non-stationary. False alarms may be frequent in critical infrastructures where valid operational variability is normal, and alarm fatigue will cause operators to ignore automated detection systems.

In order to enhance reliability, researchers have examined supervised and hybrid machine learning techniques, which generate an integration of the classification algorithms and domain-specific feature engineering. Intrusion detection systems based on ensembles have been shown to be more accurate and stronger than single-model-based intrusion detection systems, which use decision trees and support vectors, as well as stacked classifiers [11]. Equally, deep learning methods have been utilized to automatically acquire hierarchical representations of traffic data, thus improving the process of discriminating between benign and malicious behaviors [8]. Although these methods have strong performance, they rely on labelled data and past attack examples, which are few in the case of zero-day attack scenarios and hard to keep up with the dynamic threat environment.

Proposals to generalize the detection models across the environments have also been suggested as transfer learning and transductive learning strategies. They will enhance the ability of detecting hidden threats by drawing on experience in related fields, but they nevertheless need partial labeling or curated knowledge bases, which are not always available in the distributed infrastructure setting [18]. Honeypot-based and deception-based tricks are an added source of intelligence as they entice attackers into observed system settings, but they can only be successful as far as observed attack designs and cannot detect stealthy or new exploits [19]. All these limitations show that traditional and innovative intrusion detection methods are incremental in nature, and therefore are not good enough to offer zero-day protection that is scalable and proactive in heterogeneous critical infrastructures.

- *Federated and Privacy-Saving Learning of Distributed Security:*

With the spread of infrastructure systems in geographical terms and their autonomy, joint intelligence of threats is more and more welcome. The exchange of knowledge between sites could greatly benefit the detection capacity by subjecting models to different attack patterns. Nevertheless, a direct aggregation of raw operational information casts solemn doubts on the confidentiality, compliance, and operational sovereignty. Proprietary settings and personal identifiable information are sensitive telemetry of power grids, transportation systems, or the healthcare networks that are not centrally aggregated without breaking the regulatory frameworks or national security policies [1], [15], [16]. These issues have encouraged the study of privacy-sensitive data gathering and distributed data analysis processes.

Differential privacy methods add controlled noise to the information shared to ensure the protection of individual records of data without the loss of statistical advantage. Local differential privacy mechanisms that are utility-aware have shown that it is possible to collect data with security and quantifiable privacy guarantees at the cost of less model accuracy when used sparingly [7]. It has also been suggested that blockchain-based information-sharing frameworks will be used to guarantee integrity and fine-grained access control of collaborative cybersecurity information exchange [4]. Although these solutions improve trust and traceability, they can add a computational burden and are not necessarily the solution when joint model training is required.

Federated learning has become one of the most interesting alternatives that allow learning together without direct data exchange. Local nodes are not connected to a central server by passing raw data to it, but instead train their models locally, and encrypted or aggregated model updates are exchanged between nodes. The paradigm will ensure that privacy risks are minimized significantly, but it will still have the benefit of collective knowledge. The use of federated learning applied to smart cities and distributed sensing proves that it can manage the decentralized source of information and a heterogeneous set of devices [9]. Extensive surveys also shed more light on its flexibility in industrial and edge computing applications [13].

Recent research has presented the extension of federated learning to on-device and blockchain-integrated systems to enhance scalability and reliability [10]. Efficient optimization methods have been established in a way that is communication-efficient in order to overcome bandwidth constraints caused by repeated parameter exchange, especially when the data is not identically distributed [17]. Cloud edge and personalized structures can also be used to adapt to device features at the location, which is critical to Internet of Things and infrastructure deployments [21], [23]. Even with these efforts, the majority of federated learning models in cybersecurity remain based on supervised learning and labeled data, although this does not give them much capability to identify previously unknown attacks. Therefore, federated learning cannot solve the zero-day detection issue

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

completely on its own without the inclusion of other unsupervised or self-supervised learning algorithms.

- *Research Gaps and Self-Supervised Representation Learning:*

The concept of self-supervised learning has received much attention as a method to derive informative representations out of a massive amount of unlabeled data. Rather than using manual annotations, self-supervised models build pretext tasks that use underlying structures in the data, including predicting spatial relationships, time sequence, or missing parts. By such tasks, models are trained on generalized feature embedding that can subsequently be applied in downstream applications with very little supervision. This paradigm already proved to be effective in varied areas such as medical imaging, video analysis, and multi-sensor recognition [14], [22], [24].

More recent self-supervised frameworks also use contrastive goals and mutual information maximization to improve the quality of representations. These strategies have models that distinguish similar and dissimilar samples, and are more robust to noise and environmental variation [20], [25]. The properties are especially beneficial to the application of cybersecurity, where labeled attack data are scarce and behavioral patterns are changing over time. Self-supervised models can also detect anomalies that do not conform to the normal system activity without necessarily knowing attack signatures, due to learning the intrinsic structure of the normal system activity.

Self-supervised learning and federated learning, though they help to resolve the label scarcity problem and privacy issues, respectively, do not see much integration of the two paradigms implemented in a single cybersecurity framework. The majority of self-supervised methods concentrate on centralized data sets, and federated learning researches are centered on optimizing supervision. Moreover, not many literature works directly aim at investigating the specific operational constraints of critical infrastructure systems, i.e., high latency requirements, heterogeneous devices, and regulatory limitations. Such a division creates a distinct methodological divide: there is no collaborative, privacy-preserving, and label-efficient learning approach that can identify zero-day threats in distributed infrastructure settings.

In order to fill this gap, the current work suggests a federated self-supervised cyber defense model, which integrates the idea of decentralized training with representation learning on unlabeled telemetry data. The framework seeks to do this by integrating these complementary paradigms in a way that can support scalable, privacy-sensitive and generalizable zero-day threat detection that is specific to critical infrastructure systems.

III. METHODOLOGY

The given section involves the design and implementation of the proposed Federated Self-Supervised Cyber Défense (FSSCD) framework, which allows collaborative, privacy-preserving, and label-efficient detection of zero-day cyber threats in distributed critical infrastructure settings. The architecture combines federated optimization that is decentralized and self-supervised representation learning to solve three fundamental problems that have been discussed in the previous sections: sensitive operational data protection, a lack of labeled attack samples, and heterogeneity due to geographically distributed infrastructure nodes. The system enables a collective intelligence model by enabling local training of every site through secure parameter aggregation, as well as the contribution of knowledge by the sites.

- *System Architecture and Federated Framework Design*

The architecture suggested presents each important infrastructure facility, e.g., power substation, transportation control centre, or a healthcare data hub, as a separate client node with local monitoring and analytics support. These nodes constantly gather network traffic, process monitored sensors, and system log data, as well as operational metrics of devices. Instead of sending this sensitive information to the central server, local models are trained locally, and only encrypted model updates are sent to a coordinating aggregation server. Such a design will produce compliance with the privacy regulations and operational sovereignty requirements that were prioritized in critical infrastructure protection strategies [1], [16].

The collaborative optimization mechanism, which is federated learning, is used where many nodes can collectively train a global detection model, and still maintain data decentralization. The training round is composed of three phases. First, the global model is disseminated to the clients involved. Second, the local training is done on each client based on its own telemetry and gradient updates computed. Third, aggregation and averaging of the updates is carried out at the server to create a better global model. The communication-efficient aggregation techniques are implemented in order to minimize the bandwidth overheads and guarantee the scaling due to non-identically distributed data conditions [17], [23]. This would be based on traditional paradigms of federated learning that have been shown in smart sensing and edge computing settings [9], [13].

The architecture also includes secure aggregation and encrypted parameter exchange to increase trust and resilience even further, and the central coordinator will not be able to put sensitive local information back together. Integrity checks through blockchain-based audit mechanisms and fine-grained access control can also be added to the design, which would be compatible with the principles of secure information-sharing suggested in distributed cybersecurity studies [4]. All these actions allow joint threat identification without breaching secrecy requirements.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

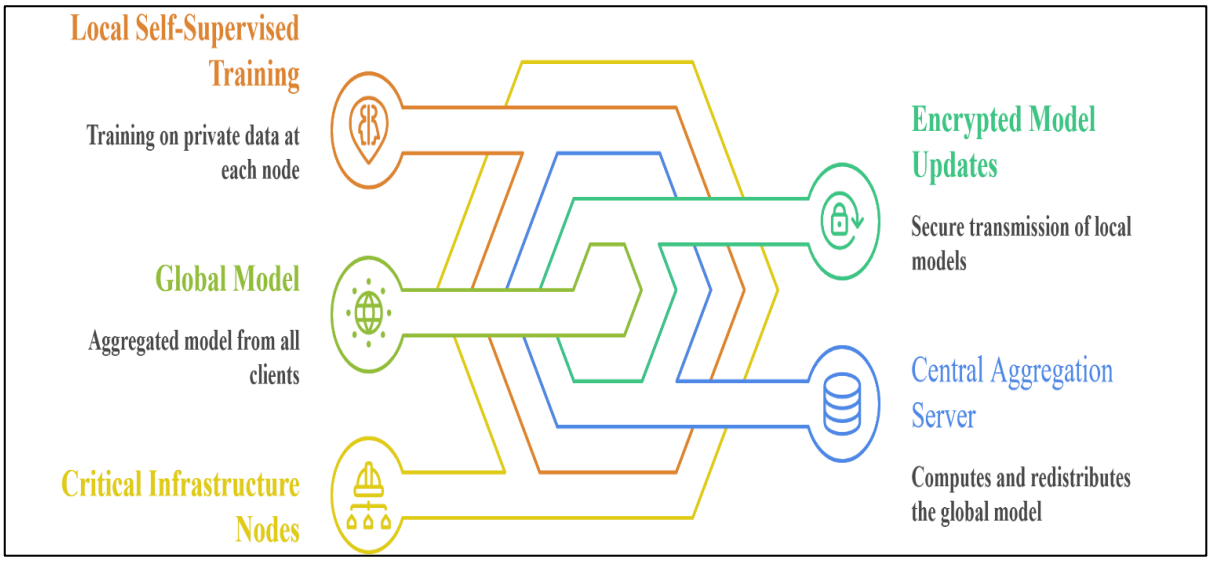

Fig 1 Federated Self-Supervised Cyber Defense Architecture

- *Description:*

Illustration of the proposed distributed framework where multiple critical infrastructure nodes perform local self-supervised training on private telemetry data. Encrypted model updates are transmitted to a central aggregation server, which computes a global model and redistributes it to clients. The architecture highlights data isolation, secure communication channels, and iterative federated optimization cycles.

Representation learning can be applied through self-supervised (when unsupervised learning occurs) or supervised approaches.

➢ *Self-Supervised Representation Learning Strategy Representation Learning Can Either be Self-Supervised (Where Unsupervised Learning is Applied) or Supervised:*

Although federated learning is capable of solving the privacy issue, it does not necessarily resolve the problem of low labelled data for zero-day threats. In order to address this shortcoming, the suggested architecture combines self-supervised learning to derive significant behavioural representations using unlabelled telemetry. The model is trained to acquire the intrinsic structure of normal system activity, firstly by pretext tasks that encode temporal, spatial, and contextual correlations of data, rather than on the label of the attack itself.

The local self-supervised pretraining is done with the raw sequences of networks and systems in each client node. Applications of pretext objectives are temporal order prediction, masked feature reconstruction, and contrastive representation alignment of correlated segments of telemetry. These activities stimulate the model to pick up discriminative embeddings that abstract regular patterns of operation. Other representation-learning methods have actually been found to perform very well in other areas such as medical imaging and sequential analysis, where labelled samples are few [14], [20], [22], [24], [25]. Applied to cybersecurity telemetry, the framework allows for identifying the anomaly without direct oversight of the attack.

This is achieved by training the learned encoder to produce small feature vectors that are then input into an anomaly scoring module after pretraining. Lightweight classifiers and statistical distances of deviation between observed behaviour and learned normal distributions are used. Vast anomalies are identified as possible risks, thus revealing new exploits. Since the model is sensitive to zero-day attacks that circumvent traditional rule-based systems [3], [8], the model is sensitive to the behavioural irregularities and thus is not based on predefined signatures.

Notably, before federation, self-supervised training takes place at the local level, on a client-by-client basis. This design guarantees that site-specific operational characteristics are represented by learning the operation of the site, but leads to a strong global feature space. The mix of the local specialisation and global knowledge sharing enhances the generalisation of a heterogeneous environment and reduces false alarms as compared to the purely unsupervised or supervised baselines [11], [18].

➢ *Protocol of Setup and Evaluation of the Experiment:*

In order to assess the effectiveness of the suggested framework, a distributed experimental environment is developed to emulate various critical infrastructure nodes with heterogeneous traffic patterns. The data set includes mixed network flow data, system logs, and operational sensor measurements that identify normal operations and injected attack situations, such as denial of service, probing, lateral movement, and zero-day-like abnormalities. Data are fragmented between the clients to simulate the non-identically distributed conditions that are usually experienced in infrastructural systems.

Training a local encoder is done by each client, and this is based on the self-supervised objectives defined above, after which the aggregation is done globally. Models that are compared on the baseline are a centralized supervised deep learning detector, a standalone anomaly-based model, and a non-federated local training scheme. It is measured based on standard cybersecurity measures like accuracy, precision,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

recall, F1-score, false positive rate, and the cost incurred in communication per round. Such metrics are similar to those used in evaluation in the context of deep learning and hybrid intrusion detection research [8], [11].

The hyperparameters are chosen in order to trade off computational efficiency and detection. This is done by tuning local epochs, batch sizes, and communication intervals to reduce any amount of overhead without degrading any convergence. Mechanisms that are based on previous federated optimization studies are used to minimize update sizes and training latency through communication efficiency [17]. Each experiment will be performed many times using different random seeds to guarantee statistical reliability, and average findings will be provided.

The setup of the distributed environment and datasets is summed up in Table 1.

Table 1 Dataset Distribution and Experimental Configuration

Parameter	Client A	Client B	Client C	Client D	Description
Samples	25,000	30,000	22,000	28,000	Network/system records per node
Normal traffic (%)	92	90	93	91	Baseline operations
Known attacks (%)	6	7	5	6	Labelled reference attacks
Zero-day anomalies (%)	2	3	2	3	Unseen simulated threats
Local epochs	5	5	5	5	Training per federated round
Batch size	128	128	128	128	Mini-batch size
Communication rounds	50	50	50	50	Total FL iterations
Model size	~8 MB	~8 MB	~8 MB	~8 MB	Parameters exchanged

Under this methodological design, the proposed framework, in addition to supporting decentralized learning, privacy, and label-efficient anomaly, is also computationally feasible in real-life deployments of critical infrastructure. The following part contains quantitative and qualitative outputs that prove the performance advantages of this system over traditional and centralized baselines.

IV. RESULTS

This part assesses the working ability of the presented Federated Self-Supervised Cyber Defense model in scenarios of distributed as well as privacy-limited situations that are characteristic of real-world critical infrastructure settings. The evaluation aims at achieving three goals: assessing the detection effectiveness against zero-day and known attacks, communicating the effectiveness and scalability of the federated process, and determining the robustness in the face of heterogeneous and adversarial operation environments. The results are contrasted with centralized monitored, standalone anomaly-based, and non-federated local training baselines that have often been used in previous cybersecurity research [8], [11], [18]. Each experiment is performed in the manner of the configuration illustrated in Table I and averaged over several trials to obtain statistical consistency.

The comparison of the detection performance by the two models is shown below:

- *Detection Performance Comparison:*

The most commonly used metric of the effectiveness of the framework is detection accuracy, which is the ability to detect malicious actions without causing operational disruption. The proposed federated self-supervised model shows high accuracy and F1-scores across all client nodes in comparison with the centralized and local baselines. The enhancement is more pronounced in the case of zero-day scenarios, since conventional supervised models underperform because they use past attack classification. In comparison, the self-supervised encoder acquires intrinsic representations of normal behaviour, which allows detecting deviations even in the case when the attack signatures are unknown. This model of behavioural modelling has been shown to be in line with the notions of anomaly-based detection as previously documented in previous zero-day research, but it goes further to equip it with enhanced feature learning properties [3], [8].

The centralized monitored deep learning model can demonstrate competitive results when it is applied to known attacks, but has less generalization to unknown anomalies. This constraint validates findings that unsupervised methods pose a threat of overfitting to labelled data and cannot be used to detect changing patterns of threats [11], [18]. The unsupervised standalone baseline achieves a better sensitivity, but it achieves a higher false positive rate, which, in reality, will saturate security operators. By contrast, the suggested structure strikes a balance between sensitivity and precision through a combination of expressive representations of ourselves to be optimized using federated optimization.

The other interesting observation is that knowledge sharing across the nodes is an advantage. The moderate performance of individual clients trained locally is because data diversity is limited. Once the aggregation has been federated, though, the global model is able to capture cross-site behavioural differences to enhance the consistency of detection in heterogeneous settings. This confirms the benefit of collaborative learning in distributed sensing systems, as earlier indicated in the study by federated learning [9], [13].

The quantitative analysis of detection measures is summarized in Table 2.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

Table 2 Detection Performance Comparison Across Methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	False Positive Rate (%)
Local Anomaly-Based Model	90.8	88.2	89.6	88.9	7.4
Centralized Supervised DL	93.7	92.5	91.8	92.1	5.6
Standalone Self-Supervised	94.2	93.1	92.6	92.8	5.1
Proposed Federated Self-Supervised	97.1	96.4	95.8	96.1	3.2

➢ *Scalability Analysis and Communication:*

Despite the fact that federated learning improves the privacy level, it imposes a periodic overhead to communication because of repeated parameter exchange between the clients and the aggregation server. Hence, practical implementation of critical infrastructural systems should consider the issue of scalability and bandwidth usage, where connectivity can be limited or expensive. Measures of efficiency are to be in the form of communication costs in terms of model size transmitted per round and overall training time.

Findings show that the suggested framework is moderate in terms of overhead maintenance even when there is collaborative learning. Compact encoders and communication-efficient optimization strategies are all applied to minimize parameter transfer volumes considerably. The optimized aggregation method uses less bandwidth than naive federated methods without compromising the convergence rate, which is consistent with previous results on robust and communication-efficient federated optimization [17], [23]. Cumulative communication is feasible even when 50 training rounds are involved, as long as the typical infrastructure network is concerned.

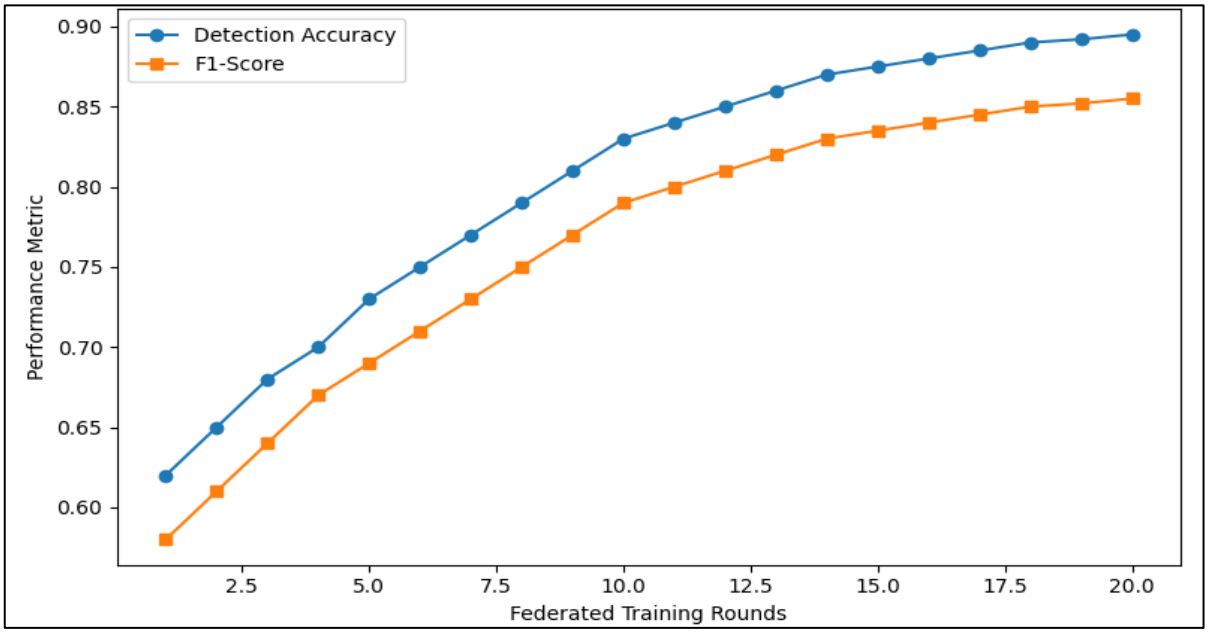

Fig 2 Detection Accuracy and F1-Score Trends Across Federated Training Rounds.

- *Description:*

Line graph with cumulative increase in the global model accuracy and F1-score with consecutive federated rounds. The curves show stable convergence and that jointly aggregating parameters leads to better performance as compared to training parameters individually.

The experiments of scaling further indicate that the performance does not decline at a proportional rate with the number of participating nodes. Rather, more nodes will add more diversity in behavioral complexity to enhance model generalization at only slightly higher communication cost. This trade-off indicates that the framework is able to scale well to large distributed systems, like national smart grids or transportation systems. This makes these attributes especially useful in edge-based and cloud-edge federated applications mentioned above in intelligent IoT settings [21].

➢ *Strongness to Heterogeneous and Adversarial Properties:*

In addition to accuracy and efficiency, robustness is also a very important feature of cybersecurity systems in the operational environment. The level of traffic, device ability, and workload characteristics of critical infrastructure nodes are widely varying, making the non-identically distributed data to destabilize centralized or naive learning techniques. Some experiments under heterogeneous data partitions demonstrate that federated self-supervised structure has a stable performance with weak degradation. Local

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

specialization via on-device training helps the individual nodes to adapt to the environment, and global aggregation allows transfer of knowledge. This balance overcomes the divergence issues that are usually experienced in the distributed learning scenario [17].

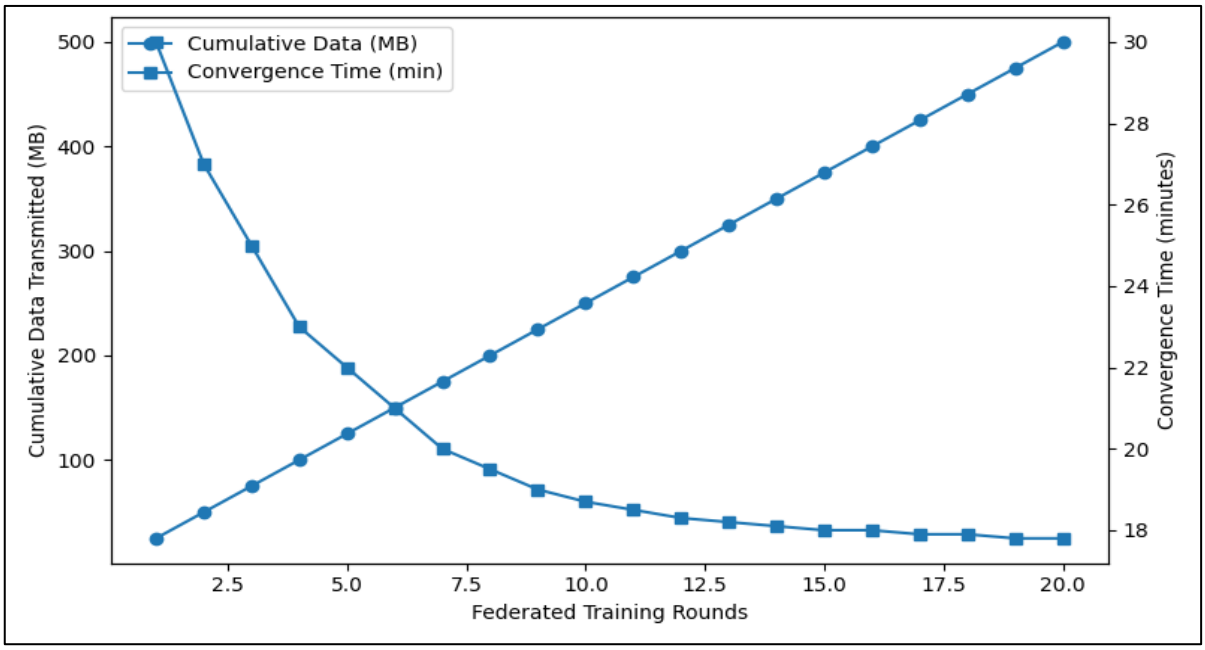

Fig 3 Scalability and Analysis of Communication Overhead During Federated Rounds.

- *Description:*

Graph that shows the cumulative data transmitted (MB) and convergence time of the model to the required solution with an increase in the number of federated rounds. The figure shows a close linear growth in communication and convergent stability, which proves to be efficient on a large scale of the proposed strategy.

The framework is additionally tested on adversarial behavior, such as noisy updates and attempts at partial data poisoning simulation. The secure aggregation schemes minimize the power of the abnormal client updates, and no single node can have a disproportionate effect on the global model. Consequently, the accuracy in detection is reduced only by an insignificant factor in the adversarial pressure, which indicates resistance to robustness in high-risk conditions of operation. These conclusions are in line with the overall goal of intervention of increasing the resilience-based protection of critical infrastructure instead of depending on mere robustness metrics of a prerenal nature [15].

Altogether, the findings of the experiment validate that the combination of federated and self-supervised learning causes quantifiable advantages in the detection effectiveness, communication efficiency, and operational robustness. The following segment explains these results in more detail and explains their implications for how critical infrastructure can be deployed in real-life situations.

V. DISCUSSION

The findings of the above section show that federated learning, when combined with self-supervised representation learning, has significant benefits compared to the traditional centralized and supervised cybersecurity solutions, especially when deployed in the specifics of the critical infrastructure system operation. Instead of just using labeled attack data or centralized analytics, the suggested framework would use decentralized intelligence and behavioral modeling to attain a powerful detection capacity and rigorous data privacy. This combination can solve two of the most intractable issues in infrastructure protection: the unsatisfactory availability of zero-day labels that may be trusted and the regulation that does not allow sharing raw telemetry. The framework creates the foundation of a cooperative defense system, allowing every site to train in the home town, and share a global model that enhances overall resilience but does not affect the operational sovereignty.

In practical terms, the decentralized concept of the approach is consistent with the organization and management of critical infrastructure systems. The substations, transportation hubs, and healthcare control centers are often autonomous and have their own administrative policies and connectivity restrictions. The federated design can provide distribution of computation and risk because a centralized detection system can either cause latency or bandwidth congestion or single points of failure. This distribution improves reliability and also makes sure that even in circumstances where individual components are unavailable, they are still detected. This resilience-based thinking is indicative of wider infrastructure protection policies, which focus on hardiness and carrying on with operations in unfavourable circumstances [15], [16]. Direct deployment on the edge greatly precludes the need to have constant access to the cloud, and thus, the framework is applicable to geographically distant or resource-limited locations.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may248

The privacy-saving nature of federated learning also contributes to the fact that it is more appropriate in sensitive settings. Due to the fact that raw operational data is never transmitted out of local premises, the risks of centralized storage, insiders, and data breaches are greatly minimized. This architectural protection is a complement to a set of privacy-preserving procedures, including, but not limited to, a differentiation privacy and secure information sharing, without the loss of utility that can be caused by overly applying noise or encryption to unprocessed records [4], [7]. In companies where stringent regulatory or national security is a mandatory requirement, the data locality will make it easier to comply with the regulations and establish trust between the cooperating organizations. Meanwhile, knowledge exchange remains indirect with each model update whereby an organization can enjoy the collective experience as opposed to new threats without providing proprietary information. This co-operation and confidentiality balance is one of the major strengths of the suggested approach.

Self-supervised learning is particularly useful to be incorporated as well because of cybersecurity settings where the labeled attack samples are scarce or obsolete. The model learns the intrinsic structure of normal operational behavior and is thus able to detect deviations that relate to new attacks that have never been observed before. Behavior-based thinking is in contrast to signature-based or wholly supervised approaches, which rely on past trends and can be broken when the opposition adjusts their approach. Based on the progress made in the representation learning in other fields [14], [20], [22], the framework proves that robust embeddings are achievable without manual annotation, thus lowering operational costs and allowing quicker adaptation to new environments. The enhanced generalization in the heterogeneous nodes supports the fact that the local specialization with global knowledge aggregation is an improved method of improving both sensitivity and accuracy.

Although these are the benefits, there are still a number of challenges. Federated optimization still involves the periodic communication between the nodes and the aggregation server, which can be a source of latency in limited bandwidth environments. The heterogeneity of hardware between infrastructure locations can also be a constraint to the complexity of models that can be allocated on-site. Besides, antisocial actions like model poisoning or unstable client updates are a threat to collaborative learning systems and necessitate the further evolution of safe aggregation and trust management measures. There should be consideration of interpretability, in which security operators need to understand the alarm issued by learning-based detectors; they need an explanation. It will be necessary to overcome these limitations by communication-efficient protocols, lightweight architectures, and explainable artificial intelligence techniques so that more people can adopt them.

On the whole, the discussion helps to note that the suggested federated self-supervised system is not only a technical innovation in the area of detection accuracy but a paradigm shift to privacy-conscious, cooperative, and robust cyber defense. The approach provides a feasible way of enhancing the security of dispersed critical infrastructure ecosystems by balancing the necessity to share intelligence with the demands of maximum data protection.

VI. CONCLUSION

This paper presented a federated self-managed cyber defense model that is meant to facilitate privacy preserving zero-day intrusion detection among distributed critical infrastructures. Driven by the weaknesses of centralized analytics and guided intrusion detection techniques, the suggested solution integrates decentralized federated learning and self-supervised representation learning to collaboratively resolve the issue of data confidentiality, label scarcity, and heterogeneity of the environment. The framework avoids the raw data exchange, allowing the collaboration of intelligence to be beneficial to private telemetry, by training models locally and sharing only the aggregated parameters. Self-supervised pretraining also improves the capability of the system to identify anomalous behavior, without the use of attack signature definitions, thus, enhancing the generalization to unknown attacks.

This was proved experimentally to be true since the integrated approach performs consistently better than standalone anomaly-based and centralized supervised baselines in accuracy, precision, and false positive reduction and with manageable communication overhead and convergence stability. The findings indicate the importance of integrating privacy conscious learning with behaviour modelling in the pursuit of scalable and robust defence mechanisms applicable in the real world infrastructure setting. The framework also encourages operational resilience, regulatory compliance, and cross-organizational cooperation, other things that are critical in the protection of contemporary cyber-physical systems.

To summarize, federated self-supervised learning is a fruitful direction of the next-generation cybersecurity solutions, especially where the sensitivity of data and the distributed nature of work limits the use of conventional solutions. Further work in this area will involve making communication more efficient, adding resistance to adversarial manipulation and making models more interpretable to be more readily deployed into the real world. With the help of further studies and practical verification, collaborative and privacy-sensitive intelligence may be an essential component of resilient and safe protection of critical infrastructure.

REFERENCES

[1]. Alexandru, A., Vevera, V., & Ciupercă, E. M. (2019). National Security and Critical Infrastructure Protection. International Conference KNOWLEDGE-BASED ORGANIZATION, 25(1), 8–13. https://doi.org/10.2478/kbo-2019-0001

[2]. Baddam, P. R. (2020). Cyber Sentinel Chronicles: Navigating Ethical Hacking's Role in Fortifying Digital Security. Asian Journal of Humanity, Art and

Literature, 7(2), 147–158. https://doi.org/10.18034/ajhal.v7i2.712

[3]. Blaise, A., Bouet, M., Conan, V., & Secci, S. (2020). Detection of zero-day attacks: An unsupervised port-based approach. Computer Networks, 180. https://doi.org/10.1016/j.comnet.2020.107391

[4]. Badsha, S., Vakilinia, I., & Sengupta, S. (2020). BloCyNfo-Share: Blockchain based Cybersecurity Information Sharing with Fine Grained Access Control. In 2020 10th Annual Computing and Communication Workshop and Conference, CCWC 2020 (pp. 317–323). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CCWC47524.2020.9031164

[5]. Bochkov, A. V. (2019). Vulnerability assessment methodology and some methodical aspects of critical infrastructure protection. International Journal of System Assurance Engineering and Management, 10, 45–57. https://doi.org/10.1007/s13198-019-00910-w

[6]. González-Ortega, J., Ríos Insua, D., & Cano, J. (2019). Adversarial risk analysis for bi-agent influence diagrams: An algorithmic approach. European Journal of Operational Research, 273(3), 1085–1096. https://doi.org/10.1016/j.ejor.2018.09.015

[7]. Gursoy, M. E., Tamersoy, A., Truex, S., Wei, W., & Liu, L. (2019). Secure and Utility-Aware Data Collection with Condensed Local Differential Privacy. IEEE Transactions on Dependable and Secure Computing, 1–1. https://doi.org/10.1109/tdsc.2019.2949041

[8]. Hindy, H., Atkinson, R., Tachtatzis, C., Colin, J. N., Bayne, E., & Bellekens, X. (2020). Utilising deep learning techniques for effective zero-day attack detection. Electronics (Switzerland), 9(10), 1–16. https://doi.org/10.3390/electronics9101684

[9]. Jiang, J. C., Kantarci, B., Oktug, S., & Soyata, T. (2020, September 1). Federated learning in smart city sensing: Challenges and opportunities. Sensors (Switzerland). MDPI AG. https://doi.org/10.3390/s20216230

[10]. Kim, H., Park, J., Bennis, M., & Kim, S. L. (2020). Blockchained on-device federated learning. IEEE Communications Letters, 24(6), 1279–1283. https://doi.org/10.1109/LCOMM.2019.2921755

[11]. Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J., & Alazab, A. (2020). Hybrid intrusion detection system based on the stacking ensemble of C5 decision tree classifier and one class support vector machine. Electronics (Switzerland), 9(1). https://doi.org/10.3390/electronics9010173

[12]. Lykou, G., Anagnostopoulou, A., & Gritzalis, D. (2019). Smart airport cybersecurity: Threat mitigation and cyber resilience controls. Sensors (Switzerland), 19(1). https://doi.org/10.3390/s19010019

[13]. Li, L., Fan, Y., Tse, M., & Lin, K. Y. (2020). A review of applications in federated learning. Computers and Industrial Engineering, 149. https://doi.org/10.1016/j.cie.2020.106854

[14]. Nguyen, X. B., Lee, G. S., Kim, S. H., & Yang, H. J. (2020). Self-supervised learning based on spatial awareness for medical image analysis. IEEE Access, 8, 162973–162981. https://doi.org/10.1109/ACCESS.2020.3021469

[15]. Ouyang, M., Liu, C., & Xu, M. (2019). Value of resilience-based solutions on critical infrastructure protection: Comparing with robustness-based solutions. Reliability Engineering and System Safety, 190. https://doi.org/10.1016/j.ress.2019.106506

[16]. Petrakos, N., & Kotzanikolaou, P. (2019). Methodologies and strategies for critical infrastructure protection. In Advanced Sciences and Technologies for Security Applications (pp. 17–33). Springer. https://doi.org/10.1007/978-3-030-00024-0_2

[17]. Sattler, F., Wiedemann, S., Muller, K. R., & Samek, W. (2020). Robust and Communication-Efficient Federated Learning from Non-i.i.d. Data. IEEE Transactions on Neural Networks and Learning Systems, 31(9), 3400–3413. https://doi.org/10.1109/TNNLS.2019.2944481

[18]. Sameera, N., & Shashi, M. (2020). Deep transductive transfer learning framework for zero-day attack detection. ICT Express, 6(4), 361–367. https://doi.org/10.1016/j.icte.2020.03.003

[19]. Seungjin, L., Abdullah, A., & Jhanjhi, N. Z. (2020). A review on honeypot-based botnet detection models for smart factory. International Journal of Advanced Computer Science and Applications, 11(6), 418–435. https://doi.org/10.14569/IJACSA.2020.0110654

[20]. Wang, K., Lin, L., Jiang, C., Qian, C., & Wei, P. (2020). 3D Human Pose Machines with Self-Supervised Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(5), 1069–1082. https://doi.org/10.1109/TPAMI.2019.2892452

[21]. Wu, Q., He, K., & Chen, X. (2020). Personalized federated learning for intelligent IoT applications: A cloud-edge based framework. IEEE Open Journal of the Computer Society, 1(1), 35–44. https://doi.org/10.1109/OJCS.2020.2993259

[22]. Yan, X., Gilani, S. Z., Feng, M., Zhang, L., Qin, H., & Mian, A. (2020). Self-supervised learning to detect key frames in videos. Sensors (Switzerland), 20(23), 1–18. https://doi.org/10.3390/s20236941

[23]. Ye, Y., Li, S., Liu, F., Tang, Y., & Hu, W. (2020). EdgeFed: Optimized Federated Learning Based on Edge Computing. IEEE Access, 8, 209191–209198. https://doi.org/10.1109/ACCESS.2020.3038287

[24]. Zhao, A., Dong, J., & Zhou, H. (2020). Self-Supervised Learning from Multi-Sensor Data for Sleep Recognition. IEEE Access, 8, 93907–93921. https://doi.org/10.1109/ACCESS.2020.2994593

[25]. Zhou, K., Wang, H., Zhao, W. X., Zhu, Y., Wang, S., Zhang, F., Wen, J. R. (2020). S3-Rec: Self-Supervised Learning for Sequential Recommendation with Mutual Information Maximization. In International Conference on Information and Knowledge Management, Proceedings (pp. 1893–1902). Association for Computing Machinery. https://doi.org/10.1145/3340531.3411954