**Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection and Implementation of Homomorphic encryption for Privacy-preserving.**

**Submitted by:**

Shagun Bala

Peiding Wang

Adnan Chowdhry

Bernard Boakye

Jamie Renouf

**For the School of Information and communication technology of Griffith University.**

**Submitted to:**

Dr. Zahra Jadidi

# **Abstract**

To protect the Cyber Infrastructure against cyber-attacks, cyber security professionals are proactively making their efforts to design a system that would be used effectively to mitigate the emerging problem of Cyber-attacks. Cyber Threat Intelligence sharing scheme is one of the cyber defense systems, which is used by cyber defenders to mitigate the problem of accelerating cyber-attacks. Cyber threat intelligence (CTI) sharing scheme proposes a system where the information about the newest threats and vulnerabilities are shared among the stakeholders to create situational awareness and eventually, remedies are implemented [1]. The machine learning (ML)-based Intrusion detection system has proven to be more efficient in the system where all the data samples belong to the same organization [2]. However, to implement CTI sharing in an ML-based Intrusion detection system, heterogeneous data samples from diverse sources are used, which is quite a challenging procedure to implement. The reason for the difficulty in its implementation is the lack of a common data format and the privacy of the organization’s data. In this paper, we propose a system based on CTI sharing where multiple organizations can share their data for the purpose of ML-based Intrusion detection, without sharing their sensitive data for privacy concerns along with the availability of common network data format of the participating organisations. In addition to these datasets, two more datasets are considered for evaluation purposes. Hence, the results present the proposed framework where ML (Machine Learning) based intrusion detection system utilizes the data from the different organisations in a universal data format to detect the intrusions in the network traffic of organisations, without being exchanging the data internally among the organisations. *(Will be updated at the completion of the project).*

# **Introduction**

Cyber threat intelligence (CTI), as explained by David Sutton, can be described as the practice of sharing, analyzing, and collecting valuable information about possible dangerous cyber threats [6]. CTI is utilized by many organizations to ensure the security of valuable information is kept safe by informing risk assessments and recognizing vulnerabilities in the systems. CTI can take numerous forms. For example, it can be vulnerabilities within both hardware and software or even data leakage about malware [1]. CTI may also be used with automated platforms that can analyze and collect significant amounts of data from numerous sources to try and detect and identify trends and patterns in cyber-attacks [6]. Therefore, CTI is particularly important for companies as there are many sources of data that can negatively affect organizations, and the public's everyday lives in the context of cyber threat security. For example: Data records of mobile phone calls, social networks, credit cards and PayPal and passport scanners [6]. Not only is this an invasion of privacy, but people and companies may also suffer monetary loss from hackers due to inadequate CTI.

CTI Sharing is a critical aspect of Cyber Threat defense systems to protect against the growing issue of potential cyber intruders and attacks. CTI sharing can be defined as the sharing of information about the most recent vulnerabilities and threats amongst stakeholders within an organization to provide awareness and allow remediation measures to be made [6]. Machine learning based intrusion detection systems are also effective to detect potential intrusion. However, utilizing CTI sharing in an ML-based IDS using heterogeneous data samples from numerous organizations can be difficult because of privacy concerns [6]. In our proposed system, the system being based on CTI sharing will allow numerous organizations to safely share their data for ML-based IDS reasons whilst avoiding privacy concerns of data. The proposed framework is evaluated by utilizing 4 datasets that will show the system's effectiveness on intrusion detection in network traffic without requiring the exchange of sensitive data amongst organizations.

Machine learning and especially federated learning enhance CTI in many ways. For example, Detection anomalies where Machine learning models are effectively trained with copious amounts of data to find anomalous patterns indicating a potential cyber-attack [5]. Additionally, another example would be a ‘collaborative threat intelligence’ where federated learning enables many organizations to collaborate on cyber threat intelligence sharing, and this is done without needing to share raw data [3]. In short, machine learning and federative learning can significantly enhance CTI via leveraging collective data and relevant knowledge from numerous organizations whilst also ensuring the protection of security and privacy of private sensitive information.

Traditional CTI approaches from organizations in the past have often just relied on internal data and analysis to detect security threats. The risk with this is of course the quality and quantity of readily available data being significantly limited to each organization, and therefore resulting in unreliable and inaccurate assessments of potential cyber threats [4]. However, this is where machine learning and federative learning are extremely useful. With federative learning, all organizations can collaborate on the development and training machine learning models via utilizing combined datasets and maintaining data security. Therefore, this allows more accurate CTI models that respond better to security threats and are more reliable to detect them early.

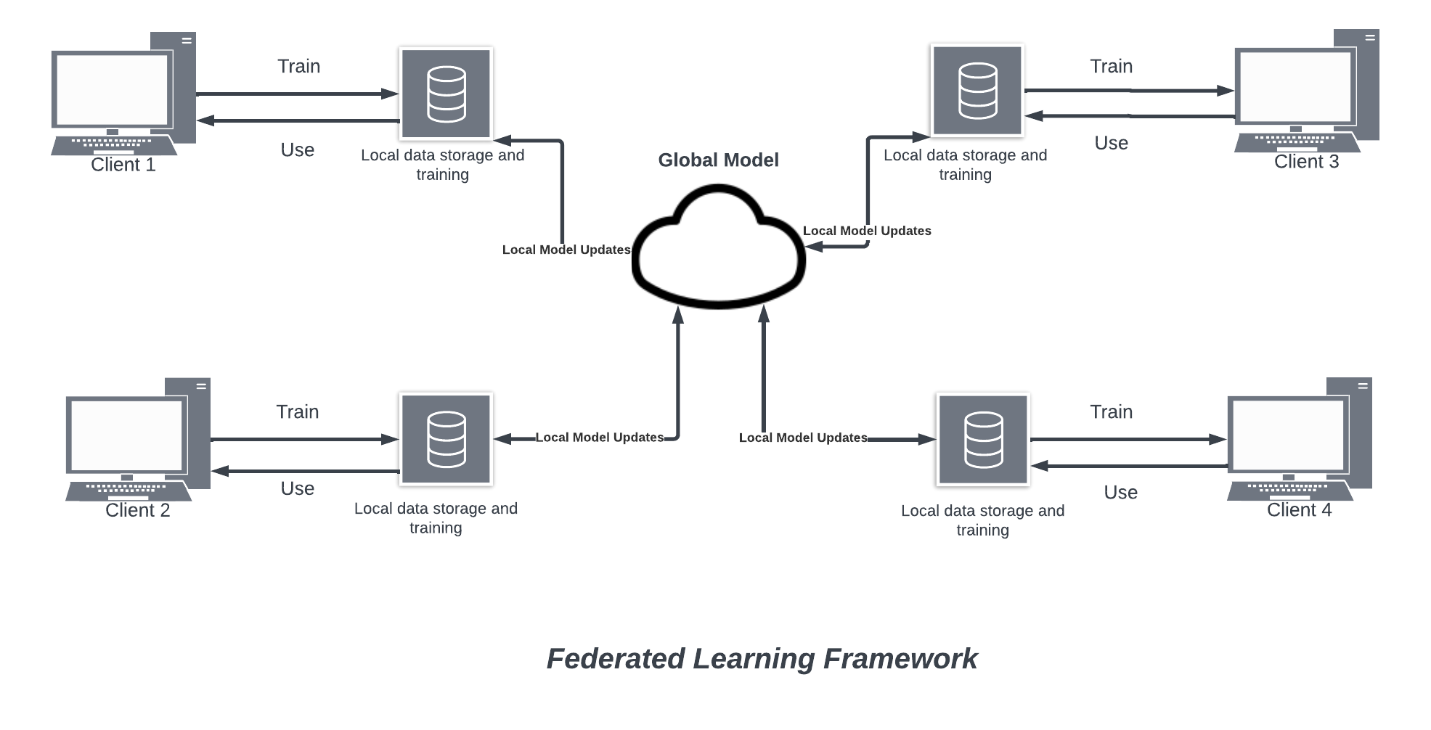
However, it is also noticed in some models that there are some limitations in Federated learning model. One of the limitations is the information leakage, when local models are aggregated into the center. So, to overcome this limitation of FL, Homomorphic encryption is the possible solution. Homomorphic encryption (HE) is the technology that allows arithmetic operations on ciphertexts without decryption. HE technique assists to protect the information of the local parties in the model. With the assistance of HE, a centralized server can make updates in the parameters of global model with the help of encrypted local gradients based on homomorphic operation [41]. Therefore, clients in the HE-based FL model would not have any risk in relation to the data leakage through local gradients, and the client delivers the encryption-based local gradients to the server.

This paper explores the application of Federated Learning in Cyber Security, and explains the experimental results conducted by the authors to set up a Federated Learning system that will enable various parties to be able to collaboratively train and work with a shared model without sharing sensitive data. Moreover, the privacy preservation of the FL model will be addressed through Homomorphic encryption. Additionally, the authors have conducted thorough research of related works and fully acknowledged the potential repercussions of implementing the federated learning framework such ASAS lack of expertise, data heterogeneity and limited resources. The authors point out gaps in the work of Yang, et. Al in [7], which is about how the Federated Learning framework (FedAI) had been developed by a team at Tsinghjua University in China. They found that the article lacks discussion of potential conflicts, challenges, and limitations of implementing a secure federated learning framework.

# **Federated Learning**

From the last few years, it has been noticed that Artificial Intelligence and the machine learning has a profound influence in the field of Information Technology. However, with the adoption of Machine Learning in various sectors such as commerce, Finance, healthcare, smart home applications, etc., there is a great amount of data that needs to be aggregated. Therefore, the privacy and security of the data is a crucial aspect to protect the data from any kind of security threats. To overcome the limitation of security issues, Federated learning emerges as a solution.

Federated Learning is a novel technology which aims to establish a collaborative Machine Learning (ML) model, in which the data is located at diverse locations [12]. Federated Learning (FL) is a privacy protection technique in which the terminal device's computing resources are efficiently used to train the Machine Learning Model. This process assists in prevention of the leakage of confidential data at the event of transmission [13]. Federated learning involves two processes: model training and model interference. In the training model, the information is exchanged among the participating parties, without exchanging sensitive data. However, model interference is executed in case of an instance of new data [12]. In other words, it is a collaborative mechanism in which the parameters from the local models of different organisations are exchanged, rather than shared their local data. The basic framework of Federated learning is depicted in *Figure 1*:



*Figure 1: Framework of Federated Learning.*

Federated learning involves four processes, in which the information of the different organisation is transferred to the server. The foremost step in FL is executing the client selection technique in which the clients are selected either by random selection from the pool or through an algorithm. Furthermore, the model’s parameter broadcasting to the selected clients is implemented. In addition to this, the participating organisations retain their models by utilizing their respective local data [14]. Eventually, at the last process, clients or organisations transfer their local model’s parameters to the global server and hence, all the aggregated parameters transfer to the global model [14].

To build a Machine learning model, different Federated Learning models are used, as per the data or expectations of the clients or organisations. One of the FL Framework is *Horizontal FL* in which the different clients of multiple organisations have the identical form of data, which is collaboratively trained on a Machine Learning model with the assistance of a cloud server or local model’s parameters of clients. There is an assumption in this framework that there is a trustworthy server and there are precise edge devices, which eventually prohibit the any kind of data loss [15]. Another approach is *Vertical FL*, in which two diverse organisations train a ML (Machine Learning) model jointly and have their own diverse data which is non-identical. For the concerns in relation to privacy and confidentiality, third party is involved in this framework, and it is assumed that the third-party is highly trustworthy. This third-party role can be played by the Government Authorities or by the secure computing node like Intel Software Guard Extension [16]. This kind of framework usually consists of two processes: Encrypted Entity Alignment and Encrypted Model Training [17]. Encrypted Entity Alignment in which common users of both the organisations are identified without actually exposing their sensitive information through the User ID Alignment techniques based on Encryption. On the other hand, Encrypted Model Training in which the identified common users’ data is utilized to train the ML model following the procedure of Federated Learning. Besides this framework, another architecture of the FL is Federated transfer learning who is quite like the traditional Machine learning, in which a new characteristic is added to the pre-trained model. It is the extension to vertical federated learning, which eventually makes some changes in the collaborative results of the participating organisations.

Currently, various attempts and explorations have been executed in real life to apply federated learning in various sectors related to industry engineering or computer science. Application of Federated learning for Mobile phones has become an important task for researchers since Google’s prediction of input by the users from Gboard on Android devices. Nowadays, users prefer a mobile device with a large storage capacity and high computing power. However, it is quite difficult to fulfil this demand because of the limitation of communication bandwidth. The solution in terms of Federated learning is addressed by combining it with Mobile edge computing (MEC), in which an ‘In-AI-framework’ is applied. In this framework, FL, based on deep reinforcement learning, collaborates with MEC system, and addresses the problem of resource allocation [18]. However, another attempt has been made in which a scheme based on privacy-aware service placement has been developed by catching the desired service on the edge server close to the users, to give the high-quality service [19]. Another approach regarding smart home IOT has also been developed, in which a multi-task learning framework enables to learn the behavior patterns of users and identify the physical hazards [20]. However, due to the feature of data privacy protection, FL is also widely used in industrial engineering. In the task of visual inspection, FL based model assists to solve the problem of detecting defects in production tasks and provides the privacy feature to the manufacturers [20]. FL is also proven to be suitable for the detection of malicious attacks in Unmanned Aerial Vehicles composed communication system [21]. In the case of electric vehicles, FL enables the prediction of charging stations for the prevention of energy congestion in transmission process [22]. Another feature of FL is seen in the field of Finance, in which FL helps the banks to identify credit card frauds [23]. Moreover, in another sector in which the FL has a profound influence is the Healthcare system. It enables us to record a large volume of heterogeneous data and examine or analyze the patients from distinct locations. To maintain the electronic health records, FL assists to collaborate the patients having similar clinical data in a record using the patient’s hashing framework and helps to predict the hospitalization in the event of cardiac issues, death or ICU stay period using cluster primal splitting FL Algorithm [23]. However, in Biomedical Imaging Analysis, `FL application algorithm is used to extract features from MRI (Magnetic resonance images) from diverse sources [24]. Thus, it can be said that the FL has a major influence in the field of smart devices, industrial engineering, and Health system. FL enables the expansion of the scope of data applications and assists in the improvement of performance of model by having collaborative operations among different entities.

Besides the extensive features of FL, there are some challenges in adopting the Federated learning setting. One of the critical problems in Federated network is the expensive communication. In federated learning, Communication includes transferring raw data or any necessary data generated in each local model. As the FL model includes ample devices who shares many data, therefore FL network should be communication-efficient to send small messages or system updates as a part of the training process, rather than transferring the entire dataset at a time. To make communication efficient, decreasing the number of rounds of communication or size of the transmitted messages are enforced. Another major challenge in enforcing FL framework is the capability to work with heterogeneous data. The heterogeneity in the network and non-similar distributed data affect the performance of the model. Fl model is still not capable enough to tackle the heterogeneous systems in network. On the other hand, Privacy is another concern in FL model. FL helps to protect the data generated from each device by exchanging models. However, exchanging model updates can lead to the exposure of sensitive information to third-party [25]. Hence, FL still needs to be improved to overcome these challenges.

# **Application of Federated Learning in Cybersecurity**

Cyber Security refers to the collaboration of security concepts, security policies, risk management strategies, training, assurance, and practices, that can be used to safeguard the cyber environment, organisations, and cyber assets [31]. The main task of Cyber security is to ensure that the organisation is secured from the security risks in the cyber environment. The main objectives of Cyber security are:

* Availability: Data should be available.
* Integrity: Access to the data should be authenticated.
* Confidentiality: To keep the data secure. (ITU, 2008)

The objective of keeping security and privacy of data, while transferring or exchanging between the different parties, is one of the challenging approaches. However, to keep the data secure in the sharing process, Federated Learning is the possible solution. FL enables the organisation to achieve their cyber-security objectives, which includes Availability, confidentiality, Integrity, data privacy. In Cyber security, Federated learning has a huge benefit and the area of applications. Some of the applications of federated Learning are as follows:

## FL in Network Intrusion Detection

In the Network Intrusion Detection system, FL helps to train the model of intrusion detection deep learning by various participating parties, without sharing any kind of network traffic. FL can be implemented in any kind of network traffic, such as Internet service providers, medical sector, Education, and any financial institutions. Regarding the different Internet service providers traffic networks, FL can collaborate with different ISPs to train a deep learning mode, without exchanging their network traffic with each other. However, about the other sectors such as medical care or Education, FL deep learning model works more effectively as their data size is smaller. In the FL deep learning model, the server is deployed in the cloud and all the participating parties connect each other with server. There are various contributions of Fl in Network Intrusion Detection system, which are as follows:

* To improve the accuracy of the detection of network attacks, it allows the collaboration of multiple organisations to train the model, by protecting their own privacy of network traffic.
* If the FL detection algorithm is proposed with the network traffic data processing method, it would assist to improve the algorithm and method of processing, in collaboration with the features of network traffic.
* As compared to the other models, the accuracy of the FL trained model is higher in network Intrusion detection. [32]

In the actual application of Federated Learning on Intrusion detection method, when each of the participating party in a model collects the network traffic, it processes the data and performs iterative training of the model. After the process of iterations, the cloud server is updated with the parameters in the model. In addition to this, the network traffic data of the clients are stored locally, whereas the models and the parameters are shared among themselves. Regarding the training data of each of the clients, it is also stored and processed locally. Hence, in this way, the authorization and the control of the data is with the participating organisations itself, which promotes the higher security and privacy of the network traffic. Another flexibility in implementing FL is that clients can withdraw their participation anytime from the model.

## FL in Cyber Threat Hunting

Federated Learning based Cyber threat hunting models is one of the recent advances in cyber security. FL based Cyber Threat Hunting assists in combating privacy issues of transferring the data into a single machine. This kind of proposed model proposes to hunt the attack samples without sharing data. It has been seen that FL based threat hunting is widely being used in the Blockchain as well as IIoT (Industrial Internet of Things) environments. Some of their respective explanations are as follows:

### *FL based Cyber Threat Hunting model in IIoT Environment*

### The FL based Cyber threat model can be used in the IIoT Environment, by initially passing the current state of the client in the model in a test phase. In the test phase, the data is sent to the model of threat sample using the global components and further sent to the classifier repository to hunt the threats. This kind of proposed framework shows high performance in terms of scalability and stability in handling diverse kinds of clients, as well as demonstrating faster training than centralized methods. In the IIoT environment, this framework is usually employed in order to detect previously seen attacks, known attacks and to hunt the previous unseen attacks. This framework works more effectively by adding cyber-threat projection system, which can predict the steps of the attackers. [33]

### *FL based Cyber Threat Hunting model in Blockchain.*

Detection of the anomalies in the blockchain based in IIoT environment, is one of the challenges. It is because, in Blockchain, each block needs to be sent to the central server, which eventually leads to an increment in training time and new block data in the testing phase. An approved approach to this kind of complexity is FL, which allows the collaboration among all the edge devices, which all the data stays on the device. In FL, after having information about each of the smart factory’s data, devices and service providers, the parameters of the model are sent to the server to update the model [34]. In Blockchain, each of the participants in smart factory, can provide a fake block or transaction to deliver a message. However, through FL, smart factories’ data and chain of blocks can be collected and shared among other local ML models. The specific data that can be used to detect anomalies are sensitive data of smart factories, characterizes of the previous forks and the amount and nature of the occurred malicious transactions. This could help to hunt the threats in the blockchain network, and this model only shares the parameters of the trained models, which can help to preserve privacy as well [35].

## FL based cyber threat intelligence sharing.

Cyber Threat intelligence sharing is gaining a tremendous importance as it enables the cyber defenders to mitigate increment in cyber-attacks in a more collaborative approach. Cyber Threat intelligence provides information about cyber-attacks, such as cyber attacker Intelligence, technical intelligence, and device log files [36]. However, in cyber threat intelligence sharing, data on various corporations are shared, which could possess a risk to the sensitive information of the companies.to solve this issue, federated learning is the suitable approach [30]. The importance of adopting Federated learning in Cyber threat intelligence are as follows:

* Federated learning can give the solution to the problem of privacy preservation in the process of cyber threat intelligence sharing. This FL based framework trains the machine learning model and enables the learning in heterogeneous decentralized databases without sharing any sensitive data.
* Because of the reduction in the transmission of data, there is less requirement of latency, power, and storage as result of a no central entity that stores all the samples of the data [37].
* In the context of Network Intrusion detection system, FL can assist in establishing machine learning models efficiently, as it includes the heterogeneous data originating from diverse sources and focuses on the privacy of network users [38].

Hence, federated learning is gaining importance in the Cyber security sector effectively. It helps better performance of cyber security and prevents increasing cyber-attacks.

# **Literature Review**

There have been many research papers which can be found on the concept of machine learning based cyber threat intelligence sharing by taking the approach of federated learning into consideration. However, it can be noticed that most of the papers that the network Intrusion detection has been conducted on the datasets arising from the local endpoints of the same organisation. On the other hand, some of the research papers have also focused on the heterogeneous datasets arising from different organisations to conduct the Network Intrusion detection system.

* In [8], Sarhan et. al. proposed the framework in which the approach of Federated learning has been utilized in the cyber security applications of Network Intrusion detection and Cyber threat Intelligence. In this framework, federated learning is proposed to share threat intelligence between two organisations in such a way that inter-organisational data should not be exchanged due to the privacy issues and non-identical format of datasets. This enables the increment in the exposure of Network Intrusion detection system to data networks originating from dissimilar sources with some benign traffic and scenarios carrying malicious attacks. In relation to this framework, two distinct datasets (NF-UNSW-NB15-v2 and NF-BoT-IoT-v2) are represented by two organisations, who share identical feature set of horizontal FL (Federated Learning).

For the evaluation and testing purposes, three distinct approaches of Federated (where two distinct clients and single global server are participated), Centralized (where two distinct organisations and central server are participated) and Localized learning (no collaboration between the participating organisations )are considered. The experiments on these three models are conducted using Google’s TensorFlow framework and for the purpose of obtaining detection performance, a Deep Neural Network (DNN) and Long Short-Term Memory (LSTM) parameters along with the hyperparameters have been utilized.

As a result, it is noticed that federated learning approach is superior as compared to localized and centralized approach in terms of privacy and cost effectiveness. However, in relation to federated learning approach to ML- based NIDS, there are certain limitations have been noticed in the event of extreme level of heterogeneous data. However, in this approach, a small amount of privacy is still sacrificed to make sharing data possible. Furthermore, reduced detection rate has been observed of 4.17% and 5.78% by utilizing the parameters of DNN and LSTM respectively, on the evaluation of the model from one dataset against the another.

* In [9], Ghimire et. al. conducted a survey highlighting the importance of Federated learning-based cyber security in Cyber-Physical System (CPS) or Internet of things (IOT) and the problems countered while applying this approach. In this paper, several federated learning-based articles are surveyed from which some of the features and challenges are addressed in relation to FL in IOT. In relation to the features of the FL based security systems, the primary focus of these solutions is the security model’s accuracy. Moreover, FL based security solutions also focus on the performance metrics as well. As per the highlighted importance of FL in the paper, it is observed that FL-based IDS (Intrusion Detection System) can identify the intrusions not previously experienced by the system. Moreover, Detection and prevention system based on Federated-self learning assists to detect and prevent the intrusions and the unknown attacks in IOT system.

Besides that, some serious threats in relation to Federated learning have also been highlighted in the paper. The threat to FL is Parameter poising and Reverse engineering ML attacks. These types of data can be performed by using the data of end devices or the Client’s model parameters. These attacks result in an adverse impact of revealing the privacy of user by spoofing on updates of model which is sent by the user’s device.

In addition to this, from the survey, it is observed that there are two popular datasets in relation to the Intrusion detection in cyber security, which are KDDCup99 and NSL-KDD. These datasets contain five major intrusion categories, such as normal, Denial of service, User to Root attacks, Remote to User and Probe.

However, from the survey, some challenges of adopting Federated learning in IOT are also addressed. It is observed that most of the Fl based proposed models utilize the Neural network, which increases the complexity and results in the increment of the overhead in heterogeneous IOT environments. Moreover, in relation to the IOT network, there is a limited communication bandwidth and in some of the proposed models based on FL scenario, edge server transfers the updates to central server by collecting them from end devices, which in general does not in IOT based environment because of not having suitable configuration. In addition, multiparty computation based on FL is experimented in small-scale network for privacy preservation. However, such kind of approach does not work in large networks because of hefty communication and computation burden. On the other hand, FL based approaches are only limited regarding measuring the level of heterogeneity in large scale networks, which can eventually lead to inaccurate outcomes of learning models.

* Jahromi et, al. proposes an ensemble-based deep Federated Learning Cyber Threat Intelligence (CTI) hunting style model to detect and attack samples in in industrial internet of things (IIoT environments, whilst avoiding sensitive data being potentially shared to ensure strong data privacy. The suggested hunting model contains 2 parallel federated-based components, with one of them evaluating the status of IIoT network based on only normal conditions, whereas the other component analyses and evaluates the network status via careful consideration of all potential threats. Furthermore, the proposed model is evaluated via two test cases, then compares it with works in the literature and outperforms other works in the f1-score metric.

The CTI hunting model is known to be very stable and reliable when it comes to taking on numerous different clients and the training time for the clients is also faster than the centralized models, even with the computational difficulty being the same. However, the method could see improvements if a more in-depth analysis of single feature values were to be added. The paper highlights that although current cyber security solutions that have already been developed for IT systems are effective, there are still additional specific methods required to accurately detect cyber-attacks in IIoT environments. Furthermore, the paper concludes that the proposed deep Federated Learning model has shown to be significantly effective and reliable for early detection of potential cyber-attacks in IIoT environments whilst avoiding the possibility of sharing sensitive data. Although the proposed model outperforms other works, there are some clear limitations in the paper.

The main limitation of the proposed model is the limited critical evaluation with a larger dataset. As discussed, the model is evaluated via two test cases and still outperforms other works. However, the performance is likely to differ if performed on a bigger dataset. Therefore, testing the model on more extensive datasets is needed to get a better understanding of its true effectiveness.

Another limitation is the limited amount of comparison with other federated learning models. The proposed model was compared to centralized models, and the issue is there are no comparisons with contrasting federated learning models. This makes it difficult to determine if the proposed model is definitively the best performing federated learning model for its use of detecting cyber-attacks in IIoT. Furthermore, the paper also does not thoroughly discuss computation costs and efficiency of the proposed model. We know the model’s training time has been shown to be faster than centralized with the same computational complexity, however, it is still not evident how much computation is needed to ensure the model’s stability is maintained throughout all the training process. Finally, the proposed model.

* Huong et al. proposes a distributed anomaly detection (AD) architecture named FedeX. Its use is to detect cyberthreats in Industrial Control Systems (ICSs) of smart IoT-based factories. The suggested design employs federated learning to quickly train and acquire data patterns, whilst also remaining sufficiently light to work on edge devices that have a limited number of resources available. Additionally, this paper critically evaluates FedeX via 2 datasets, specifically the liquid storage dataset, and the SWAT, and demonstrates how FedeX outperforms 14 other current anomaly detection solutions on every detection metric when applied to liquid storage dataset.

FedeX has shown itself to be extremely fast in terms of its training time and low hardware demands, and this makes it suitable for the deployment of anomaly detection applications on top of edge computing systems and achieving real-time detection. Furthermore, the paper highlights the significance of Explainable Artificial intelligence (XAI) in comprehending the identified anomalies to enable experts to make informed decisions promptly and raise their confidence and trust in the model. The presented design presents a viable and promising approach for edge computing in distributed settings. However, the article also points out the need for further investigations and research to tackle all privacy concerns when sending sensitive data over communication channels.

While the proposed FedeX architecture has displayed promising results for detection of anomalies in IoT-based ICSs, there are some limitations that must be considered. For example, limited evaluation datasets: the article assesses the effectiveness of FedeX architecture via 2 datasets, namely the SWAT and liquid storage dataset. Whilst both datasets encompass different scenarios and anomalies, the FedeX architecture will need more varied and intricate datasets to validate the effectiveness in real-world settings. Additionally, another limitation would be the limiting discussion on privacy and security: The paper has not extensively explored the potential problems of privacy and security in IoT-based ICSs, and only gives brief discussions as to how the FedeX architecture tackles the concerns. It is vital to prioritize the protection of data security in ICSs, especially when sensitive data is continuously being transmitted, and therefore, more attention is essential in this area of concern.

* Jiang et all. Presents a new method for sharing Cyber Threat Intelligence (CTI) called Blockchain and Federated Learning (FL) for sharing threat detection models as CTI called (BFLS). The main issue the paper addresses is the difficulty in sharing CTI because of privacy issues and concerns that causes many organisations to be reluctant to share sensitive information with other organisations. The proposed solution is based on blockchain technology and Federated Learning. The goal is to directly address privacy issues and concerns preventing organisations from sharing sensitive information. Fl is used to train a threat detection model and it blocks potential privacy leakages from happening. Blockchain is also used to share the model and overcome all potential risks of server failures and malicious nodes. Simply by training the model on distributed data, FL can maintain robust local learning on local devices, and users can have a trusted well-trained threat detection model without personal data leaked to the central server. According to the experimental results on CIC-DDoS-2019 and ISCX-IDS-2012 datasets, they both showed BFLS can share CTI securely and shows securely and in its threat detection ability. The accuracy rates of BFLS were recorded as 98.92% and 98.56%.

The ISCX-IDS-2012 and CIC-DDoS-2019 datasets have been used to assess performance. The ISCX-IDS-2012 dataset contains network traffic data that has several types of potential attacks such as DoS (Denial of Service), infiltration, and scanning. Whereas, CIC-DDoS-2019 dataset is man-made, and it encompasses a wide range of harmful DDoS attacks.

The limitation of the paper is it only considering FL as a solution for maintaining privacy in CTI sharing. However, although FL can protect an organisations privacy, it has been shown to be still vulnerable to adversarial malicious attacks. Additionally, FL may suffer from data imbalance problems, and this could negatively affect a model’s performance significantly. Furthermore, the proposed solution is limited to its use of its blockchain-based CTI sharing platform and might not be suitable for other CTI sharing platforms. Therefore, blockchain could have limitations such as its performance and scalability. Finally, the evaluation is limited to just 2 datasets, and the generalizability of the proposed solution to other datasets is uncertain.

* + Alazab et al. explore how Federated Learning (FL) can enhance cybersecurity and help prevent numerous different potentials cyberattacks in real-time. The authors survey various FL models developed by researchers that provide privacy, attack detection and authentication. FL enables different devices to understand and learn a collaborative machine learning model without data being shared with centralized servers, and this solves privacy concerns, increases reliability and scalability, and decreases latency. The authors also discuss real-time use cases that utilize FL to help maintain the overall privacy of data and improve system performance. The authors do this by using numerous different datasets to assess and evaluate the performance of FL models in detecting cyber-attacks of all kinds. However, the paper’s limitation is it does lack an in-depth assessment of performance for FL models for large datasets. The paper ends by identifying all key challenges and future research directions for implementing FL in real-time scenarios. The paper’s authors summarize by highlighting the vital obstacles and areas for further investigation, so that researchers can concentrate on applying Federates Learning in real-time situations. The paper's authors summarize by highlighting important obstacles and potential areas for further investigation that researchers can concentrate on to apply Federated Learning in real-time situations.

Another limitation of the paper is how it concentrates on the positive outcomes of implementing FL in cybersecurity, but it does not provide an in-depth analysis of the difficulties and challenges associated with implementing FL. For example, FL models are more complex than traditional machine learning models because they need extra stages for combining and distributing the model. This increases complexity and is likely to make it more challenging to train and adjust FL models. It could also potentially increase the risk of model overfitting. Also, FL depends on the presence of many participating devices, and this may not always be practical or achievable.

FL also faces a hurdle regarding data heterogeneity. This is where varied types and formats of data on different devices can be challenging to aggregate and process data effectively, or in meaningful way. This can lead to potential prejudiced models that have reduced reliability and accuracy. Finally, FL models are susceptible to adversarial attacks. For example, malevolent intruders insert malicious data into the FL system or interfere with the model aggregation process and undermining the reliability and accuracy of the resulting model. Overall, FL has the possibility of enhancing cybersecurity. However, it is especially important to carefully consider these limitations when creating and implementing FL-based systems.

* + Tian Li, et al. Propose a new approach for improving cyber threat intelligence (CTI) using federated learning. The typical problem with traditional CTI is how it depends on centralized data sources that can be vulnerable to several data breaches and attacks. Federated learning is a potential remedy because it enables models to be trainable on decentralized data sources, whilst preserving data privacy from being leaked.

The author's proposed method employs Federated Averaging, a type of federated learning framework which enables collaborative training of machine learning models via spread out data across numerous sources. The authors applied this framework to a dataset of cyber threat indicators containing malicious URLS, file hashes and IP addresses that are indicative to cyber threats. The evaluation was conducted on 3 datasets: The Global Threat Intelligence Feed, the Malware Domain Blocklist, and the Emerging Threats feed.

The authors discovered that their federated learning method achieved comparable levels of accuracy to traditional centralized approaches while still providing strong data privacy. However, they point out that their approach does have limitations. For example, it requires a considerable number of sources to achieve high accuracy, and the potential for data bias if sources contain different data distributions. Furthermore, the authors acknowledge that their testing was restricted to a small number of datasets, and that further research is needed to accurately determine if their approach can be used with other areas of CTI.

* Zhang et, all [39]. Proposes a system solution, BatchCrypt for Federated Learning (FL) in a cross-silo setting. In this setting, numerous organisations work together to train a machine learning model while preserving the privacy of all sensitive data. The proposed approach aims to minimize encryption and communication overhead that comes with additively homomorphic encryption (HE) used to secure each client’s update during aggregation. With BatchCrypt, a batch of quantized gradients updates is transformed into a longer integer and is encrypted at once, therefore, resulting in a significant decrease in encryption and communication overhead.

Encryption based privacy preserving techniques are needed in FL because it involves numerous organisations or devices to collaborate to train a machine learning model without raw data being shared. Encryption-based privacy preserving techniques use cryptographic algorithms to encode data in ways that can be processed without being exposed to unauthorized parties.

The paper presents novel quantization and encoding techniques and a gradient clipping approach that supports gradient-wise aggregation on ciphertexts of encoded batches. The proposed solution was implemented as a plugin module in FATE, an industrial FL framework. The evaluation was made with EC@ clients from across geographically distributed data centers which had resulted in 23 x -93 x training acceleration, as well as 66 x -101 x reduction in communication overhead, and less than 1% accuracy loss made by quantization errors. However, there are some drawbacks to the proposed approach such as the inability to handle large batch sizes, as well as the dependence on a trusted aggregator.

Additional limitations of the proposed system are it assumes the presence of a trusted aggregator that collects and aggregates encrypted gradients from numerous clients. This assumption might not work out well in all scenarios because there are likely to be cases where the aggregator is compromised, and this could lead to privacy breaches. Furthermore, security threats are another limitation of the paper, as it does not discuss potential security threats that may compromise the system. For example, side-channel attacks might leak sensitive information during computation.

* In [40], Zhang, et.al (2022), proposed a novel federated learning scheme, which focuses on the privacy-preservation in IoT-based healthcare applications consisting of diverse medical institutes. The paper emphasizes on the homomorphic encryption and the secure multi-party computing to promote the privacy preservation even in the case of collusion among honest but curious participants in the model, by deploying Diffie-Hellman key exchange and Shamir secret sharing algorithm. In addition to this, to overcome the issue of increased communication overhead, an adjustment to the EIGamal encryption algorithm is considered to transform the algorithm from multiplicative homomorphism to additive homomorphism.

In the proposed model, an optimal global model comprises of the collaboration between the servers and the clients in an epoch. Furthermore, the responsibility of some security parameters such as public keys or private keys, is addressed by a third-party trust authority.

For evaluation, “Human Against machine with 10,000 training images” (HAM10000), has been used, which has about 10,015 dermatoscopic images of pigmented skin lesions and has seven diverse classes of skin cancer. For the improvement of the accuracy of the model, the original dataset HAM10000 has been expanded and some techniques of data augmentation have been applied. The expanded dataset is further split into training set and testing set with randomness. Furthermore, the training set is distributed among each client to train the model, in the form of different subsets. The training dataset is distributed to each client by considering the security and privacy of the image and is also split as 90% for training the model and the rest of 10% for the evaluation as the validation set. Moreover, the number of parameters in the local model is 306,063 and the accuracy of the model is determined based on testing of the dataset done by the server.

As a results of the experiments in the paper, 76.9% accuracy has been achieved to detect the lesion cell type. Moreover, the variable called data quality is encrypted for each client in each training epoch to overcome the problem of increase in computation overhead. However, the proposed scheme is inefficient in heterogeneous environments to obtain high efficiency. Research on the Malicious server is also not addressed in the proposed framework.

* In [42],He, et.al (2022), proposed a FedIPEC scheme, which is based on the preservation of privacy and low latency Federated learning method, which assists to transfer the parameters that are encrypted using homomorphic encryption algorithm, for the purpose of the privacy of the data of end devices, without any transmission of the data to the nodes in the edge. An improvement or update is also made to the Paillier encryption, by performing the computation of multiple *rn mod n*2. This computation is made to modify the large exponential modular multiplication operations into the basic operations, while in the process of encryption. Time for the computation is also considered, by creating key pairs only once and perform the encryption after each local training.

The datasets used for the evaluation purposes are public data sets MNIST digit recognition, consisting of a training set of 600000 samples, having 28 x 28 grayscale handwritten digital image in each sample. For the training at each mobile device, a CNN (Convolutional Neural Network) model is used, having two convolutional layers and fully connected layers.

By the evaluations and experiments in this paper, it can be concluded that by making improvements in encryption, by introducing new hyperparameter and computing multiple at the time of the generation of key, proves to be successful in reducing the whole time required to train the model. In addition to this, FedIPEC scheme assists to efficiently protect the privacy by utilizing the Paillier encryption and the accuracy of the testing is equivalent in FedAVG (federated averaging) and the actual Paillier algorithm. However, the main challenge that is addressed in this paper is that FedIPEC scheme results in the low latency of the improved Paillier algorithm as compared to the original Paillier algorithm and chain-PPFL.

* In [43], Wibawa. et.al, proposed an FL algorithm using homomorphic encryption on real medical data to preserve privacy. The algorithm is based on the convolutional neural network (CNN) and utilizes the highly secured muti-party computational protocol to safeguard the deep learning model from any kind of attacks or theft of data.

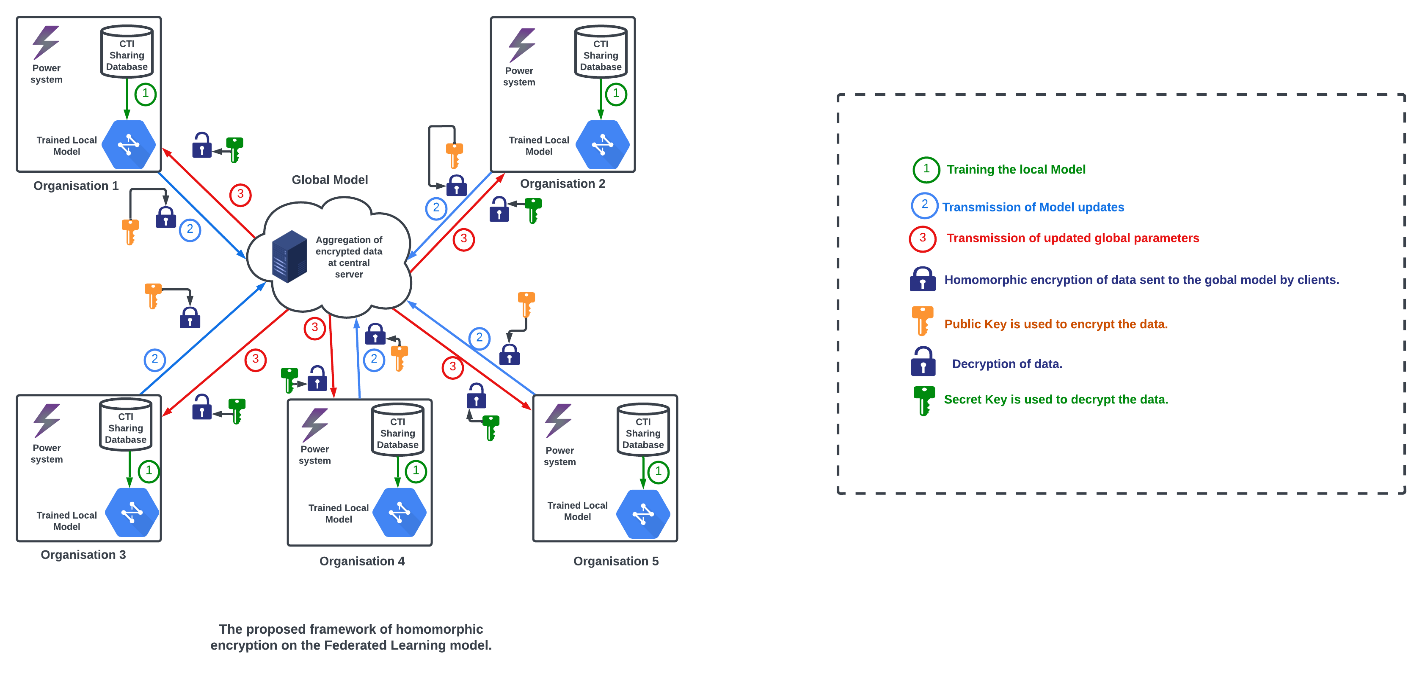
The proposed protocol and training are performed in Python, using the Keras/ TensorFlow libraries for the initial FL model and for the implementation of homomorphic encryption, Microsoft SEAL library has been used. To perform testing or examinations, a public dataset COVID-19 x-ray scans with the number of clients have been used.

After the experimental phase, it has been concluded that the overall prediction performance remains stable around 0.84 in the training model. In addition to this, the execution time of the encrypted domain is longer than the plain domain because of the complexity of the homomorphic encryption and the processing of encrypted data. However, homomorphic encryption is successfully applied, and the data is efficiently secured.   
  
The following Table shows the summary of the papers that have been reviewed in this report:

|  |  |  |
| --- | --- | --- |
| **PAPER** | **DATASETS** | **CHALLANGES** |
| **Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection.** | * NF-UNSW-NB15-v2 * NF-BoT-IoT-v2 | * A small amount of privacy is still sacrificed to make sharing data possible. * A reduced detection rate has been observed of 4.17% and 5.78% by utilizing the parameters of DNN and LSTM, respectively. |
| **Recent Advances on Federated Learning for Cybersecurity and Cybersecurity for Federated Learning for Internet of Things.** | * KDDCup99 * NSL-KDD. | * Fl based proposed models utilize the Neural network, which increases the complexity and results in the increment of the overhead in heterogeneous IOT environments. * FL based approaches are only limited regarding measuring the level of heterogeneity in large scale networks, which can eventually lead to inaccurate outcomes of learning models. |
| **An ensemble deep federated learning cyber-threat hunting model for Industrial Internet of Things.** | * Secure Water Treatment (SWAT) | * The main limitation of the proposed model is the limited critical evaluation with a larger dataset. As discussed, the model is evaluated via two test cases and still outperforms other works. * Another limitation is the limited amount of comparison with other federated learning models. The proposed model was compared to centralized models, and the issue is there are no comparisons with contrasting federated learning models. |
| **Federated Learning-Based Explainable Anomaly Detection for Industrial Control Systems.** | * liquid storage * dataset and the SWAT dataset. | * The paper has not extensively explored the potential problems of privacy and security in IoT-based ICSs, and only gives brief discussions as to how the FedeX architecture tackles the concerns. |
| **Blockchain and Federated Learning for sharing threat detection models as Cyber Threat Intelligence.** | * ISCX-IDS-2012 * CIC-DDoS-2019 | * The limitation of the paper is it only considering FL as a solution for maintaining privacy in CTI sharing. However, although FL can protect an organisations privacy, it has been shown to be still vulnerable to adversarial malicious attacks. * The evaluation is limited to just 2 datasets, and the generalizability of the proposed solution to other datasets is uncertain. |
| **Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions.** | * Data sent from end devices to the cloud in the centralized learning model approach. * Local data generated by each device in the distributed on-site learning model approach. * Local raw data generated by each client in the Federated Learning (FL) model approach. | * The limitation of the paper is how it concentrates on the positive outcomes of implementing FL in cybersecurity, but it does not provide an in-depth analysis of the difficulties and challenges associated with implementing FL * FL models are more complex than traditional machine learning models, which can make it more challenging to train and adjust FL models and increase the risk of model overfitting. |
| **Federated Learning Challenges, methods, and future directions.** | * The Global Threat Intelligence Feed * The Malware Domain Blocklist * The Emerging Threats feed | * The limitations are it requiring many sources to achieve high accuracy and the potential for data bias if sources contain different data distributions. * The authors acknowledge that their testing was restricted to a small number of datasets, and that further research is needed to accurately determine if their approach can be used with other areas of CTI. |
| **BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning.** | * BatchCrypt | * The limitations are the inability to handle large batch sizes, as well as the dependence on a trusted aggregator. * Additional limitations of the proposed system are it assumes the presence of a trusted aggregator that collects and aggregates encrypted gradients from numerous clients |
| **Homomorphic encryption-based privacy-preserving federated learning in IoT-enabled healthcare system.** | * HAM10000 | * The proposed scheme is inefficient in heterogeneous environment to obtain the high efficiency. * Research on the Malicious server is also not addressed in the proposed framework |
| **Privacy-Preserving and Low-Latency Federated Learning in Edge Computing.** | * MNIST digit recognition | * Low latency of the improved Paillier algorithm as compared to the original Paillier algorithm and chain-PPFL. |
| **Homomorphic Encryption and Federated Learning based Privacy-Preserving CNN Training: COVID-19 Detection Use-Case** | * COVID-19 x-ray | * The execution time of the encrypted domain is longer than the plain domain. |

# **Experimental Results**

## **ARCHITECTURE**



The above figure demonstrated the proposed framework of homomorphic encryption on the Federated learning model. In the framework, each client possesses its own cyber threat intelligence on power system datasets, having information about the power consumption, generation, or other relevant variables. In the first process, each of the clients trains its local model, as depicted in **1.** In the next process, before participating in the federated learning process, each of the clients encrypt its parameters using public key, by implementing the techniques of homomorphic encryption. Encryption is responsible for the assurance of the confidentiality of the data and cannot be understood or accessed by any other party, including the central server, which is coordinating the process of federated learning. The clients performed the computation, in the form of performing scaling on their encrypted weights or parameters and sent the model updates to the global server as depicted in 2. Furthermore, in the framework, the server aggregates the encrypted parameters from the clients. After performing the aggregation of encrypted parameters, the server then initiates the process of distributing the updated model to the clients for further iterations as depicted in 3, allowing the model to gradually improve the collaboration while preserving the privacy of individual client’s data. The clients will further decrypt the updated weights using the secret key. Thus, by employing homomorphic encryption, the participating clients or organisations can securely collaborate in the framework, without sharing or exposing their sensitive information to any other party.

## **DATASETS**

In order to accurately assess and determine the effectiveness of our proposed collaborate CTI sharing approach for Network Intrusion Detection System (NIDS) utilizing Federated Learning, our team conducted numerous experiments on three datasets: Binary Dataset, Three-class dataset and Multiclass dataset. All three datasets have been selected to specifically represent various network structures, potential classification and numerous attack scenarios. Below is a detailed description of each of our selected datasets:

**Binary Dataset:**

* Format: CVS (Comma-Separated Values).
* Compatibility: Compatible with Weka (open-source suite for machine learning algorithms).
* Description: it is specifically designed for network intrusion detection where the goal is to classify instances as “Attack” or “Natural”. Binary dataset has a collection of features that are essential for detecting network intrusion. For example, source and destination IP addresses, many different traffic patterns, protocols, and port numbers.

**Three-class Dataset:**

* Format: CVS.
* Compatibility: Designed for Weka.
* Description: it is specifically designed for network intrusion detections where the goal is to classify instances into three distinct categories: “Natural”, “Attack” and “no event”. Three-class dataset provides a set of features that capture many different aspects of both network traffic and certain behaviors and patterns which enable the detection of various levels of network intrusion. This dataset facilitates the analysis and identification of network anomalies across numerous threat levels, and this leads to a deeper understanding of the network security involved.

**Multiclass Dataset:**

* Format: ARFF (Attribute-Relation File Format).
* Compatibility: Designed for Weka.
* Description: it is presented in ARFF format and is suitable for handling multiclass classification tasks. It encompasses many different instances that represent diverse types of network traffic, all assigned directly to a specific class or label. The dataset consists of 37 classes in string format. The features are carefully selected within the multiclass dataset to obtain important information about network traffic patterns and protocols.

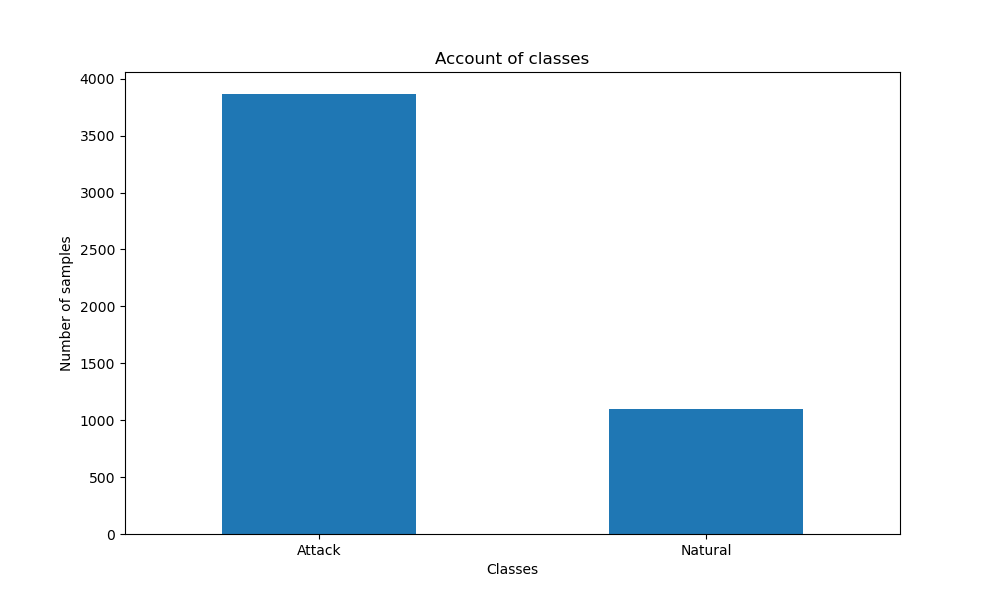
However, for the evaluation of our framework of collaborative CTI sharing scheme based on federated learning, Power System Datasets have been used. Power System Attack Datasets consists of both synchro phasor measurements and data logs from Snort, a simulated control panel, and relays related to electric transmission system normal, disturbance control and many cyber-attack behaviors. The power system dataset is further categorized into three classes, which are Binary, Multi-class, and Triple class. All the classes are made from one dataset, which has fifteen sets with 37 power system event scenarios in each. These 37 event scenarios are further categorized into natural events, no attacks, and attack events. The type of scenarios in the dataset includes Short-circuit fault, Line maintenance, Remote tripping command injection, Relay setting change attack, and Data injection attack.

* **Binary**: It is CSV format, compatible with Weka. (Will be explained more with the aid of completion of table)
* **Three-class**: It is CSV format, compatible with Weka.
* **Multiclass**: It is ARFF format, which was further converted into CSV format, that can be easily used with Weka.

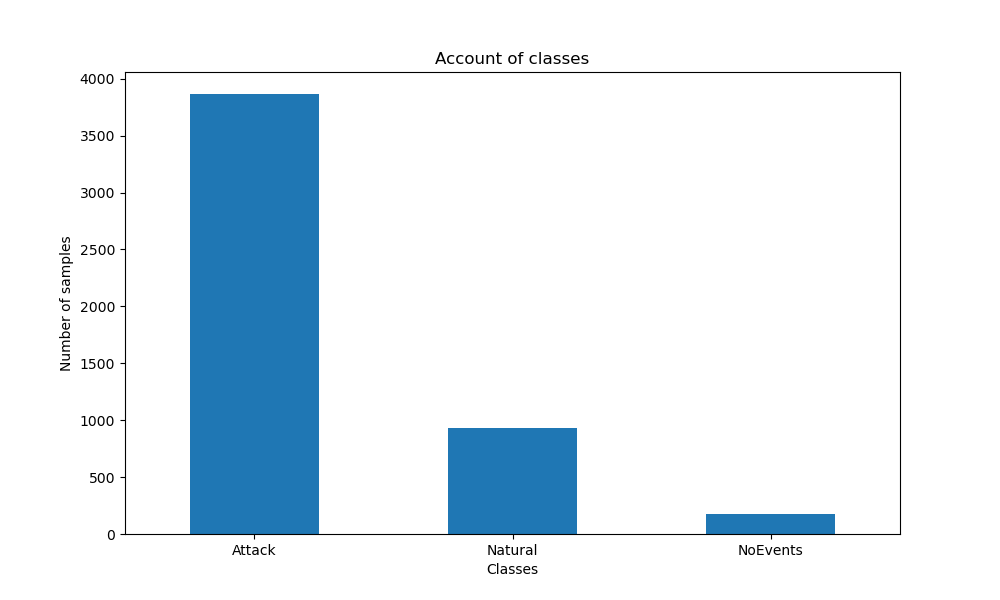
Table 1: Dataset’s target classes distribution

|  |  |  |  |
| --- | --- | --- | --- |
| **Classes of Power system Dataset** | **Binary Class** | **Multi class** | **Triple class** |
| Total samples | 4966 | 4966 | 4966 |
| Classes | 2 | 37 | 3 |

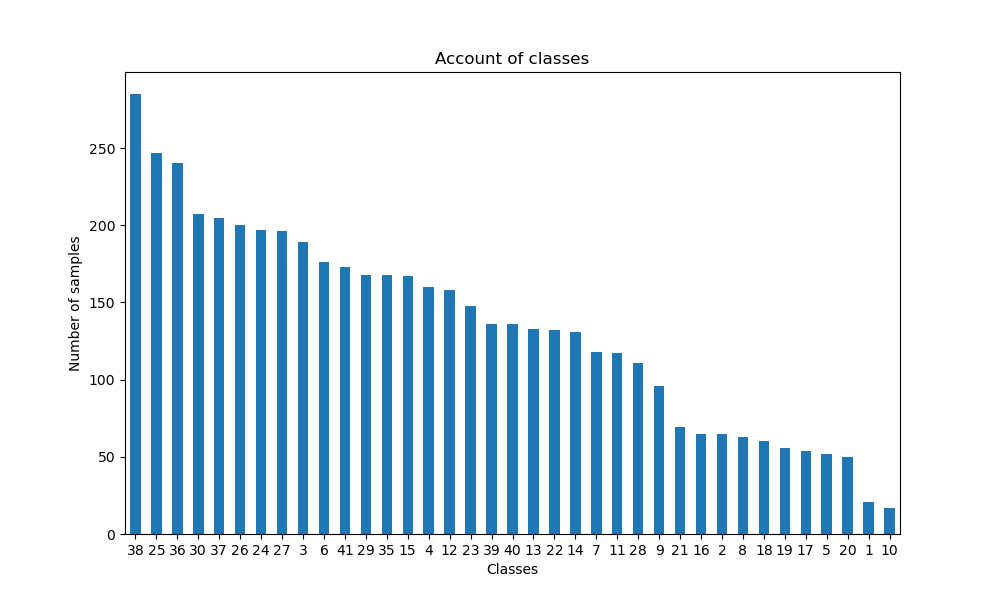
**Classes Distribution for Binary Dataset**



**Classes Distribution for Triple Dataset**



**Classes Distribution for Multi Dataset**



The above table and charts illustrate the total samples and the number of classes in each dataset categorized as Binary, Multi-class and Triple-class. Binary consists of 3866 attack and 1100 natural(benign) samples. The Multi-class dataset consists of 37 Classes which were encoded in numeric form from 0 to 36. The Triple class dataset consists of 3866 attack, 927 natural(benign), and 173 no event samples.

All three datasets contain 4966 Total samples, providing a comprehensive set of instances for assessing the proposed framework. Additionally, all datasets were utilized to evaluate the team’s proposed collaborative CTI sharing approach for NIDS via Federated Learning. They were specifically chosen to represent diverse network structures, many different attack scenarios and potential classifications. The Power System datasets have been utilized to assess the proposed framework, which is essential in data-focused applications within power systems and is very important for the development and comparison of numerous power system applications. These datasets are essential for many important tasks such as machine and deep learning techniques, novelty detection and fault analysis.

## **EXPERIMENTAL METHADOLOGY**

Figure 1 illustrates the distribution of traffic data where each organisation uses an ML model from the global server and attempts to successfully train it on its own local data samples. The updated parameter is set from each organisation, and then sent back to the global server to then calculate the averages of all the weights to produce the global model. The metrics employed in table 2 are calculated in binary format using True Positive (TP) and True Negative (TN) that show the correct number of attacks and benign data samples.

In the metrices, accuracy of the different classes and model is calculated by having a percentage of correctly classified samples from the test set. Whereas precision refers to the number of true positives divided by the total number of positive predictions. The recall measures the true positive rate, which is the number of true positives divided by all positive examples in the  
test set.

In Cyber threat intelligence, precision represents the number of times the attacks have correctly been identified. On the other hand, Recall measures how often the correct attacks are identified, which are an attack. Recall and precision are used jointly to find the F1 scores.

Table 2: Evaluation metrices

|  |  |  |
| --- | --- | --- |
| **Metric** | **Definition** | **Equation** |
| Accuracy | Percentage of correctly classified samples from the test set. |  |
| Precision | The number of true positives is divided by the total number of positive predictions. |  |
| Recall | The number of true positives divided by all positive examples in the test set. |  |
| F1 score | Harmonic means of Precision and Recall. |  |
| Support | The number of occurrences of each particular class in the true responses. | N/A |

Table 3: Training Parameters

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Binary** | **Triple** | **Multi** |
| Local epochs | 50 | 50 | 50 |
| Batch size | 64 | 64 | 64 |
| Optimizer | Adam | Adam | Adam |
| Learning rate | 0.001 | 0.001 | 0.001 |
| Number of clients | 2 | 2 | 2 |
| Loss function (Cross-entropy) | Binary | Sparse-Categorical | Sparse-Categorical |
| Federated learning rounds\* | 30 | 30 | 30 |

Table 3 represents the set of parameters of three of the different categories of datasets, that is Binary, Triple and Multi-class datasets, in the context of Federated Learning Framework. For all the three datasets, the local training epochs are set to 50, which means that each participant in the federated learning process will perform 50 epochs. In addition to this, each of the datasets has 64 samples to train the local model. The optimizer used for all the datasets is Adam, which is a well-known optimization algorithm. Moreover, the learning rate of each dataset is 0.001, which assists to control the step size in the optimization process.

Furthermore, the loss function for the binary dataset is binary cross-entropy, which is most used for the Binary classification of data. However, for the Triple and Multi-class dataset, Sparse Categorical cross-entropy is used, which is commonly used with the mutually exclusive classes. Moreover, in the federated learning model, each of the participants performs 30 iterations, to improve the global model.

Table 4: Hyperparameters for ANN (Artiﬁcial Neural Network)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Binary Class:**  **Nodes &Activation function** | **Triple Class:**  **Nodes &Activation function** | **Multi Class:**  **Nodes &Activation function** |
| **Input layer** | 89 input features | 89 input features | 89 input features |
| **Hidden layer 1** | 256 & Relu | 256 & Relu | 256 & Relu |
| **Hidden layer 2** | 128 & Relu | 128 & Relu | 128 & Relu |
| **Hidden layer 3** | 64 & Relu | 64 & Relu | 64 & Relu |
| **Output Layer** | 1 & Sigmoid | 3 & Softmax | 37 & Softmax |

The architecture and behavior of the ANN for all classes are determined from the hyperparameters. The input layer receives a total of 89 input features from the dataset out of 128 features. Several data preprocessing methods are applied to make the dataset ready for training such as dropping columns resembling target column classes, replace inf values to zero, identifying outliers and replace them with NaN value before dropping the columns which had only the outlier values. The subsequent hidden layers process all data via relu activation function shown. This function introduces non-linearity. The output layer of ANN generates the final outcomes, as well as its configuration, number of nodes and also the activation function varying that depends on the class. For instance, the multi-class dataset shows the output layer having a total of 37 nodes representing contrasting classes and uses the SoftMax activation function to produce a range of probabilities for each of the classes. The hyperparameters are essential because they define the behavior and performance of the ANN for each class, and this enables the model to quickly learn and make accurate predictions based on the input features given by the datasets

**An approach of Homomorphic encryption for the privacy-preserving of proposed Federated Learning Framework**

For the privacy preserving of our Federated learning model, homomorphic encryption was applied to the parameters of the model. We conducted a series of experiments to evaluate the performance of our encrypted Federated learning model. For the federated learning model, five participants and a central server are participating. The training process consists of multiple rounds, each round involves the training of the model at the local level and the aggregation of encrypted gradients at the server. Homomorphic encryption is employed to encrypt the model parameters or weights, which are based on the requirements of the security and the computational capabilities of the system. To evaluate the performance of the homomorphic encryption, we considered the accuracy as the performance metrices, which involves the correct prediction of digit labels for the test dataset.

## **RESULTS**

**Binary Dataset**

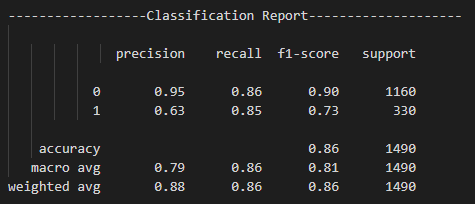
The average model training loss and accuracy for clients produced the following matrices on training data.

|  |  |  |
| --- | --- | --- |
|  | Training Loss | Training Accuracy |
| Matrices at epoch 0 | 0.6490 | 0.6186 |
| Matrices at epoch 29 | 0.0334 | 0.9885 |

The precision, recall and f1-score are important metrics to look at in classification reports to assess performance. Precision measures the accuracy of the model’s positive predictions and quantifies the amount of positive instances predicted correctly. Recall measures the model’s ability to identify the positive instances correctly and quantifies the amount of true positive instances that have been identified by the model out of every positive instance. fF1-score is a statistical measure, it combines precision and recall to one metric and provides a balanced assessment of the model’s performance. It is calculate as the harmonic mean of both recall and precision, and considers both false positives and false negatives. The classification report displays these metrics for each class and provides a comprehensive overview of how the each model performs on both individual classes, as well as an overall assessment of the effectiveness. The precision, recall and f1-score provides comprehensive insights on how the model is performing and shows areas that need improvement.

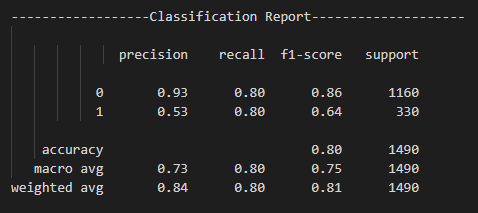
After training for 30 global rounds, both nodes have the following f1-scores on test data.

**Local\_Nodes\_1**

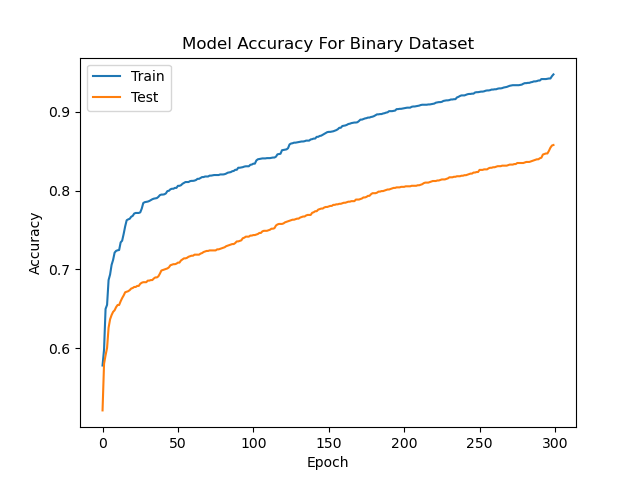


The model performs very well on class 0 with an outstanding precision score of 0.95, recall score of 0.86 and f1-score of 0.90. Class 1 however, has a decrease in performance with the precision score of 0.63, a similar recall score of 0.85 and an f1-score of 0.73.

**Local\_Nodes\_2**



The model performs very well in class 0 with an outstanding precision score of 0.93, recall score of 0.80 and f1-score of 0.86. Class 1 however, has a decrease in performance with the precision score of 0.53, a similar recall score of 0.80 and an f1-score of 0.64.



The model performs very well on class 0 with a precision score of 0.96, recall score of 0.88 and f1-score of 0.92. Class 1 however, has a decrease in performance with the precision score of 0.67, a recall score of 0.88 and an f1-score of 0.76.

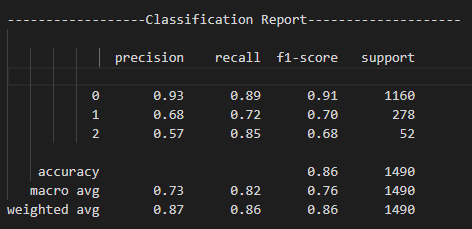
**Triple Dataset**

The average model training loss and accuracy for clients produced the following matrices on training data.

|  |  |  |
| --- | --- | --- |
|  | Training Loss | Training Accuracy |
| Matrices at epoch 0 | 0.7588 | 0.6489 |
| Matrices at epoch 29 | 0.0182 | 0.9941 |

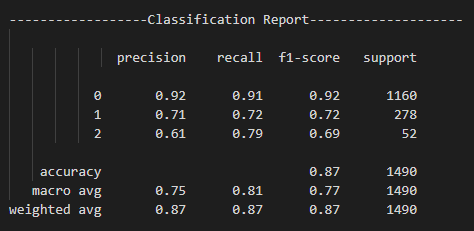
After training for 30 global rounds, both nodes have the following f1-scores on test data.

**Local\_Nodes\_1**

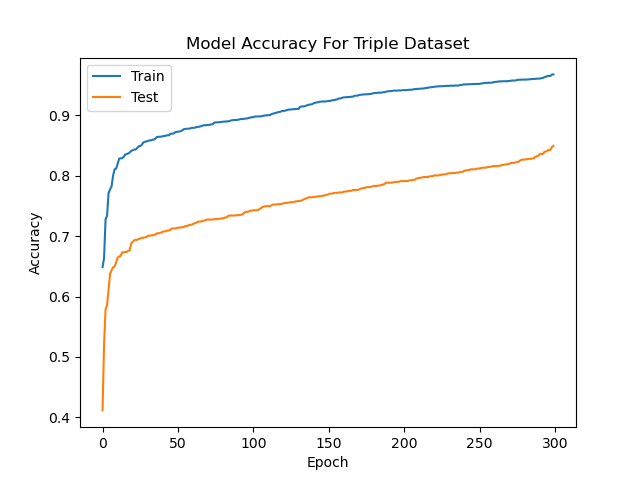


The model performs very well on class 0 with an outstanding precision score of 0.93, recall score of 0.89 and f1-score of 0.91. Class 1 however, has a decrease in performance with the precision score of 0.68, a recall score of 0.72 and an f1-score of 0.70. Class 2 also shows a decrease in performance with a precision score of 0.57, a good recall score of 0.85, and a f1-score of 0.68.

**Local\_Nodes\_2**



The model performs very well on class 0 with an outstanding precision score of 0.92, a good recall score of 0.91 and f1-score of 0.92. Class 1 however, has a decrease in performance with the precision score of 0.71, a recall score of 0.72 and an f1-score of 0.72. Class 2 also shows a decrease in performance with a precision score of 0.61, a good recall score of 0.79, and a f1-score of 0.69.



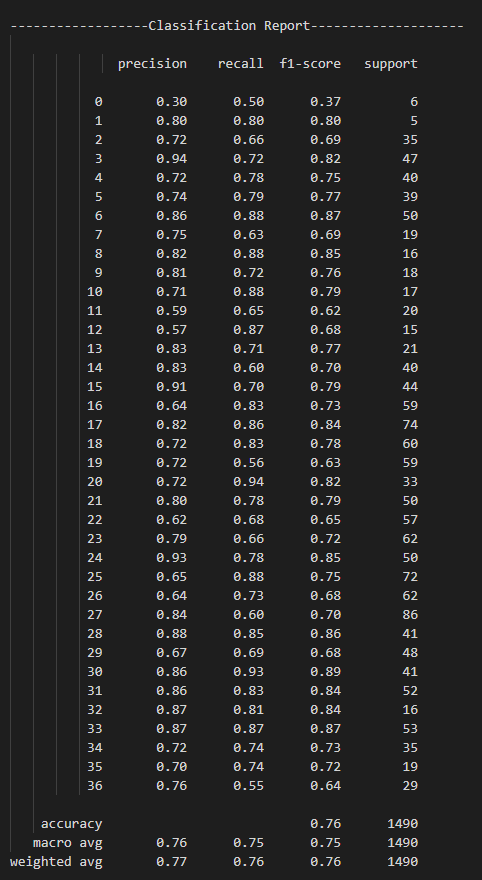
**Multi Dataset**

The average model training loss and accuracy for clients produced the following matrices on training data.

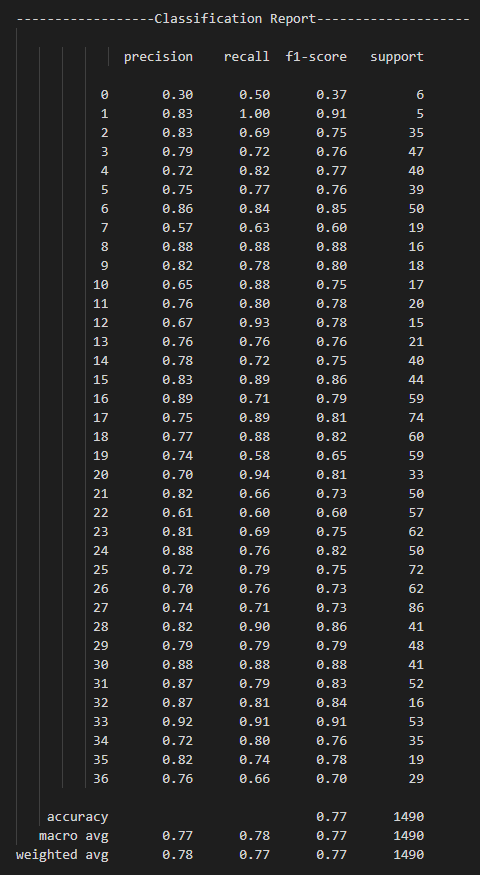
|  |  |  |
| --- | --- | --- |
|  | Training Loss | Training Accuracy |
| Matrices at epoch 0 | 3.5182 | 0.0772 |
| Matrices at epoch 29 | 0.0252 | 0.9897 |

After training for 30 global rounds, both nodes have the following f1-scores on test data.

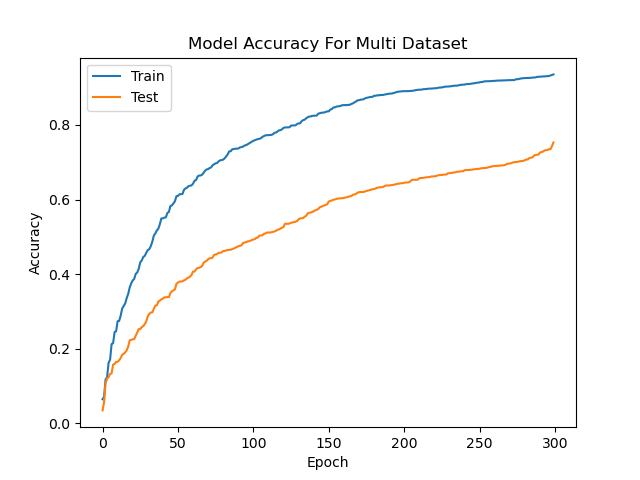
**Local\_Nodes\_1**



**Local\_Nodes\_2**



The model performs below satisfactory on class 0 with a precision score of 0.30, recall score of 0.50 and f1-score of 0.37. Class 1 however, has an improved performance with the precision score of 83, an outstanding recall score of 100 and an f1-score of 91. Class 2 shows a decrease but still a good performance with a precision score of 0.83, a recall score of 0.69, and a f1-score of 0.75.



As shown above, our train accuracy is better than test accuracy. Test accuracy is lower because our test data classification and classes are imbalanced. The graph shows where the values start to where they end, showing the whole journey of the accuracy for both training accuracy (in blue) and test accuracy (in yellow).

**Federated learning Model**

We executed our suggested methodology; we trained the model successfully and improved accuracy. Our approach proved to effectively detect many cyber-attacks in ICS, and the model’s accuracy was improved from using the available datasets. The overall performance of the model was evaluated by utilizing numerous metrics. For example, precision, recall and FI-score. Overall, the results of our report demonstrate the efficiency and effectiveness of our proposed approach in detecting cyber-attacks in ICS and highlight the significance of employing suitable datasets to improve the accuracy of the model.

The f1\_score in the classification report above is a statistical metric providing a balanced assessment of the classification model’s recall and precision performance. A higher f1\_score means better performance. It is calculated as the harmonic mean of precision and recall, and it considers the false positives and false negatives in its calculation. The precision score quantifies how often the model’s positive predictions are correct. For example, a precision score of 1 means that all positive predictions the model has made are correct, and a lower score indicates incorrect predictions have been made. The recall score measures the proportion of true positives vs actual positives. It quantifies the number of positive instances the model was able to correctly identify. For example, a recall score of 1 shows the model has successfully identified all positive instances, and a lower score indicates that there are positive instances missed. The classification report is a text-based summary of the main classification metrics and provides a convenient way to evaluate the overall performance of classification models and to assess its performance on individual classes.

**Binary-Dataset:**

* The overall accuracy of the model is 0.88, meaning that it correctly predicted 88% of the instances in the dataset.
* The macro average precision, recall, and F1-score across all three classes are 0.82, 0.88, and 0.84.
* The weighted average precision, recall, and F1-score across all three classes are 0.90, 0.88, and 0.88.

**Triple-Dataset:**

* The overall accuracy of the model is 0.86, meaning that it correctly predicted 86% of the instances in the dataset.
* The macro average precision, recall, and F1-score across all three classes are 0.73, 0.88, and 0.79.
* The weighted average precision, recall, and F1-score across all three classes are 0.82, 0.80, and 0.80.

**Multi-Dataset:**

* The overall accuracy of the model is 0.80, meaning that it correctly predicted 80% of the instances in the dataset.
* The macro average precision, recall, and F1-score across all three classes are 0.80, 0.82, and 0.80.
* The weighted average precision, recall, and F1-score across all three classes are 0.89, 0.86, and 0.87.

The model performs satisfactorily on class 0, but its performance on class 1 is inadequate. However, as the support for class 2 is low, it is challenging to arrive at firm conclusions regarding the model's performance on this class. On the whole, the model's performance is moderately good, as indicated by the weighted average F1-score.

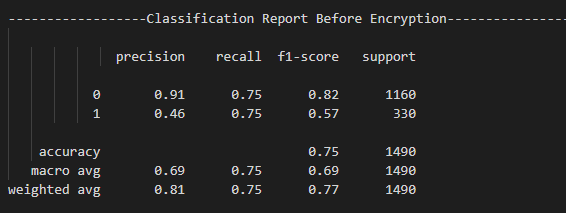
**Homomorphic Encryption**

To validate the implementation of homo-morphic encryption the following matrices are randomly extracted from the “output\_binary\_enc.txt” file provided in the GitHub repository. The matrices belong to epoch 28 of the federated learning round where the following f1-scores are produced on test data.

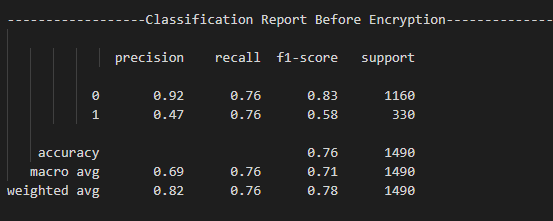
Homomorphic encryption is a cryptographic method which enables computations to be conducted on encrypted data without decrypting it first. It allows operations to be conducted directed on encrypted data whilst preserving privacy and security. For the Microsoft Tensy library we used to do homomorphic encryption. We used the public key to encrypt and private key to decrypt.

For homomorphic encryption, because the datasets and accuracy are same / very similar, the only difference is the encryption and decryption process. To validate that our encryption and decryption is not affecting the parameters, we only use the encryption to protect the privacy of the parameters as shown below:

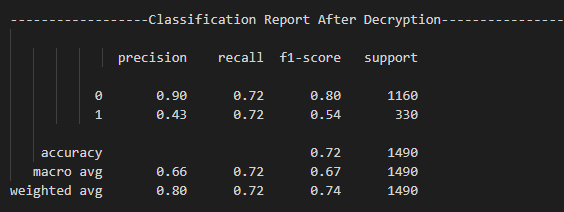
**Local\_Nodes\_1**



**Local\_Nodes\_2**



To validate the homo-morphic encryption and decryption process the following matrices are produced after decrypting the average weights and the difference between the average and training f1-scores are under the acceptable range of 3%.



From the results, we observed that the precision, recall f1-score are similar before encryption to what they are after encryption. This means the homomorphic encryption is a success because if the results were to vary significantly, it would mean the encryption has done more than what it has intended, which is to simply preserve the privacy and security of the data. The difference is only 3% and this difference is acceptable. If for example, the difference were 20% or more, then we have a problem. Therefore, our homomorphic encryption as a whole is a success.

# **Conclusion**

In conclusion, this report has presented a novel approach for cyber threat intelligence sharing, leveraging federated learning and homomorphic encryption techniques. Through the application of federated learning, network intrusion detection models can be trained collaboratively across multiple entities without the need to centralize sensitive data. The use of homomorphic encryption ensures privacy-preserving capabilities, allowing participants to securely share encrypted data and obtain useful insights without compromising sensitive information.

The results of the study demonstrate the feasibility and effectiveness of the proposed scheme. By combining the power of federated learning and homomorphic encryption, organizations and entities can collectively improve their network intrusion detection capabilities without sharing raw data or exposing confidential information to potential threats.

The implementation of the proposed scheme showcases promising performance in terms of accuracy, privacy preservation, and scalability. The federated learning rounds, with their well-defined parameters and optimization techniques, enable participants to iteratively refine and aggregate their models while protecting individual data privacy. Additionally, the use of homomorphic encryption ensures that the data remains encrypted throughout the learning process, allowing participants to share encrypted models and make predictions without decrypting the sensitive information.

Overall, the proposed cyber threat intelligence sharing scheme offers a valuable solution for organizations aiming to enhance their network intrusion detection capabilities while upholding privacy and data security. The combination of federated learning and homomorphic encryption opens new possibilities for collaborative threat intelligence sharing without compromising sensitive data, contributing to a more robust and secure cybersecurity ecosystem. Future research could focus on further optimizing the scheme's performance, exploring the applicability in real-world scenarios, and addressing any potential challenges or limitations to ensure its practical implementation and adoption.

# **References**

1. Wagner, T. D., Mahbub, K., Palomar, E., & Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers & Security*, *87*, 101589.
2. Glanz, E., Fallen, N., O'Reilly for Higher Education (Firm), & Safari, an O'Reilly Media Company. (2021). What is federated learning? (1st ed.). O'Reilly Media. Retrieved March 26, 2023, from <https://rb.gy/3ljz7m>.
3. Guerra, P., & Tamburello, P. (2018). Modernizing cybersecurity operations with machine intelligence: advanced threat detection, hunting, and analysis (First). O'Reilly Media. Retrieved April 2, 2023, from <https://proquest.safaribooksonline.com/9781492035992>.
4. Jian, J., Chen, S., Luo, X., Lee, T., & Yu, X. (2022). Organized cyber-racketeering: exploring the role of internet technology in organized cybercrime syndicates using a grounded theory approach. *Ieee Transactions on Engineering Management*, *69*(6). <https://doi.org/10.1109/TEM.2020.3002784>
5. Sarhan, M., Layeghy, S., Moustafa, N., & Portmann, M. (2022). Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection. Journal of Network and Systems Management. Advance online publication. <https://doi.org/10.1007/s10922-022-09631-1>
6. Sutton, D. (2022). Cyber security: the complete guide to cyber threats and protection (Second). BCS, The Chartered Institute for IT. Retrieved March 19, 2023, from <http://surl.li/fpngw>.
7. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: concept and applications. *Acm Transactions on Intelligent Systems and Technology (Tist)*, *10*(2), 1–19. <https://doi.org/10.1145/3298981>
8. Sarhan, M., Layeghy, S., Moustafa, N., & Portmann, M. (2023). Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection. *Journal of Network and Systems Management*, *31*(1), 3.
9. Ghimire, B., & Rawat, D. B. (2022). Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things. *IEEE Internet of Things Journal*
10. Jahromi, A. N., Karimipour, H., & Dehghantanha, A. (2023). An ensemble deep federated learning cyber-threat hunting model for Industrial Internet of Things. *Computer Communications*, *198*, 108-116.
11. Huong, T. T., Bac, T. P., Ha, K. N., Hoang, N. V., Hoang, N. X., Hung, N. T., & Tran, K. P. (2022). Federated learning-based explainable anomaly detection for industrial control systems. *IEEE Access*, *10*, 53854-53872.
12. Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T., Yu, H. (2020). Federated Learning for Vision, Language, and Recommendation. In: Federated Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer, Cham. <https://doi.org/10.1007/978-3-031-01585-4_8>
13. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, *216*, 106775.
14. Mammen, P. M. (2021). Federated learning: Opportunities and challenges. *arXiv preprint arXiv:2101.05428*.
15. Khokhar, F. A., Shah, J. H., Khan, M. A., Sharif, M., Tariq, U., & Kadry, S. (2022). A review on federated learning towards image processing. *Computers and Electrical Engineering*, *99*, 107818.
16. Aono, Y., Hayashi, T., Wang, L., & Moriai, S. (2017). Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, *13*(5), 1333-1345.
17. Bahmani, R., Barbosa, M., Brasser, F., Portela, B., Sadeghi, A. R., Scerri, G., & Warinschi, B. (2017). Secure multiparty computation from SGX. In *Financial Cryptography and Data Security: 21st International Conference, FC 2017, Sliema, Malta, April 3-7, 2017, Revised Selected Papers 21* (pp. 477-497). Springer International Publishing.
18. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1-19.
19. Qian, Y., Hu, L., Chen, J., Guan, X., Hassan, M. M., & Alelaiwi, A. (2019). Privacy-aware service placement for mobile edge computing via federated learning. *Information Sciences*, *505*, 562-570.
20. Yu, T., Li, T., Sun, Y., Nanda, S., Smith, V., Sekar, V., & Seshan, S. (2020, April). Learning context-aware policies from multiple smart homes via federated multi-task learning. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)* (pp. 104-115). IEEE.
21. Han, X., Yu, H., & Gu, H. (2019). Visual inspection with federated learning. In *Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part II 16* (pp. 52-64). Springer International Publishing.
22. Mowla, N. I., Tran, N. H., Doh, I., & Chae, K. (2019). Federated learning-based cognitive detection of jamming attack in flying ad-hoc network. *IEEE Access*, *8*, 4338-4350.
23. Saputra, Y. M., Hoang, D. T., Nguyen, D. N., Dutkiewicz, E., Mueck, M. D., & Srikanteswara, S. (2019, December). Energy demand prediction with federated learning for electric vehicle networks. In *2019 IEEE global communications conference (GLOBECOM)* (pp. 1-6). IEEE.
24. Yang, W., Zhang, Y., Ye, K., Li, L., & Xu, C. Z. (2019). Ffd: A federated learning-based method for credit card fraud detection. In *Big Data–BigData 2019: 8th International Congress, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25–30, 2019, Proceedings 8* (pp. 18-32). Springer International Publishing.
25. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60.
26. Jiang, T., Shen, G., Guo, C., Cui, Y., & Xie, B. (2023). BFLS: Blockchain and Federated Learning for sharing threat detection models as Cyber Threat Intelligence. *Computer Networks*, *224*, 109604.
27. Alazab, M., RM, S. P., Parimala, M., Maddikunta, P. K. R., Gadekallu, T. R., & Pham, Q. V. (2021). Federated learning for cybersecurity: concepts, challenges, and future directions. *IEEE Transactions on Industrial Informatics*, *18*(5), 3501-3509.
28. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60.
29. Zhang, C., Li, S., Xia, J., Wang, W., Yan, F., & Liu, Y. (2020, April). Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning. In *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC 2020)*.
30. Fereidooni, H., Dmitrienko, A., Rieger, P., Miettinen, M., Sadeghi, A. R., & Madlener, F. (2022). Fedcri: Federated mobile cyber-risk intelligence. In *Network and Distributed Systems Security (NDSS) Symposium*
31. von Solms, R., & van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, *38*, 97–102. <https://doi.org/10.1016/j.cose.2013.04.004>
32. International Telecommunications Union (ITU). ITU-TX.1205: series X: data networks, open system communications and security: telecommunication security: overview of cybersecurity 2008.
33. Tang, Z., Hu, H., & Xu, C. (2022). A federated learning method for network intrusion detection. *Concurrency and Computation: Practice and Experience*, *34*(10), e6812.
34. Jahromi, A. N., Karimipour, H., & Dehghantanha, A. (2023). An ensemble deep federated learning cyber-threat hunting model for Industrial Internet of Things. *Computer Communications*, *198*, 108-116.
35. Yazdinejad, A., Dehghantanha, A., Parizi, R. M., Hammoudeh, M., Karimipour, H., & Srivastava, G. (2022). Block hunter: Federated learning for cyber threat hunting in blockchain-based iiot networks. *IEEE Transactions on Industrial Informatics*, *18*(11), 8356-8366.
36. Jiang, T., Shen, G., Guo, C., Cui, Y., & Xie, B. (2023). BFLS: Blockchain and Federated Learning for sharing threat detection models as Cyber Threat Intelligence. *Computer Networks*, *224*, 109604.
37. Yang, K., Jiang, T., Shi, Y., & Ding, Z. (2020). Federated learning via over-the-air computation. *IEEE Transactions on Wireless Communications*, *19*(3), 2022-2035.
38. Preuveneers, D., Rimmer, V., Tsingenopoulos, I., Spooren, J., Joosen, W., & Ilie-Zudor, E. (2018). Chained anomaly detection models for federated learning: An intrusion detection case study. *Applied Sciences*, *8*(12), 2663.
39. Zhang, C., Li, S., Xia, J., Wang, W., Yan, F., & Liu, Y. (2020, April). Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning. In *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC 2020)*.
40. Zhang, L., Xu, J., Vijayakumar, P., Sharma, P. K., & Ghosh, U. (2022). Homomorphic encryption-based privacy-preserving federated learning in iot-enabled healthcare system. *IEEE Transactions on Network Science and Engineering*.
41. Park, J., & Lim, H. (2022). Privacy-preserving federated learning using homomorphic encryption. *Applied Sciences*, *12*(2), 734.
42. Park, J., & Lim, H. (2022). Privacy-preserving federated learning using homomorphic encryption. *Applied Sciences*, *12*(2), 734.
43. He, C., Liu, G., Guo, S., & Yang, Y. (2022). Privacy-Preserving and Low-Latency Federated Learning in Edge Computing. *IEEE Internet of Things Journal*, *9*(20), 20149-20159.