Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions

Mamoun Alazab, Swarna Priya RM, Parimala M, Praveen Kumar Reddy Maddikunta, Thippa Reddy Gadekallu, *Senior Member, IEEE*, and Quoc-Viet Pham, *Member, IEEE*

***Abstract*—Federated learning (FL) is a recent development in artificial intelligence, which is typically based on the concept of decentralized data. As cyberattacks are frequently happening in the various applications deployed in real time, most industrialists are hesitating to move forward in adopting the technology of the Internet of Everything. This article aims to provide an extensive study on how FL could be utilized for providing better cybersecurity and prevent various cyberattacks in real time. We present an extensive survey of the various FL models currently developed by researchers for providing authentication, privacy, trust management, and attack detection. We also discuss few real-time use cases that have been deployed recently and how FL is adopted in them for preserving privacy of data and improving the performance of the system. Based on the study, we conclude this article with some prominent challenges and future directions on which the researchers can focus for adopting FL in real-time scenarios.**

***Index Terms*—Attack detection, authentication, cyberattacks, cybersecurity, decentralized, federated learning (FL), privacy, trust management.**

I. Introduction

DURING the recent era, machine learning (ML) techniques have been successful thanks to three major factors that contributed a lot to its wide usage and success story. The first important factor is the availability of Big Data, which is gathered over various domains such as image processing and mobile networking. The second factor comes from recent advances in computational power and newer learning models. The third factor is the evolution of deep learning (DL) models that are recently used for indulging intelligence to the ML models and the computational devices. The usage of DL models has shown a very high success rate and a very good example is Alpha-Go [1] board game. Though ML has shown a large success rate, most domains are not in a position to use them in real time due to the following major reasons:

1) the users are more concerned about data privacy;
2) the confidentiality of user data is compromised since the data should be collected by the server for central learning;
3) a great amount of data and computing resources available at the network edge is not exploited effectively for learning improvement.

Although there are many companies who provide a very well-trained ML models for imparting the knowledge at a lower computational costs, there are many privacy and confidentiality issues that are not addressed yet.

To overcome the aforementioned challenges, federated learning (FL) has been proposed as a promising ML concept. Conceptually, FL enables different devices to learn a collaborative ML model without the need of data sharing with the centralized server [2]. Compared with the central learning concept, in which all devices must upload their data to the server for central storage and central learning, FL has the potential to solve privacy concerns, reduce the latency, and increase the scalability and reliability. These features are enabled by the fact that computing and learning tasks are widely distributed throughout many devices in the network, and the learning server, equipped with edge computing capabilities, is merely responsible for model aggregation and broadcasting. Thanks to its distinctive features, FL has been leveraged in many areas such as mobile and wireless networking [3], [4], healthcare [5], and vehicle communications [6]. FL has also found numerous applications in cybersecurity areas. There are various types of attacks prevailing in cybersecurity such as malware, intrusion detection, denial of service (DoS), floods of the system, and zero-day attack. For example, the conventional intrusion detection system requires data collection and storage from edge and Internet-of-Things (IoT) devices. Such centralized learning is not efficient because of various reasons such as large communication and latency cost, no willingness to share private information, and learning burden on the centralized server [7]. FL appears as a great solution

Manuscript received June 22, 2021; revised August 29, 2021; accepted October 6, 2021. Date of publication October 11, 2021; date of current version February 2, 2022. The work of Mamoun Alazab was supported by the National Research Foundation of Korea under Grant NRF-2021S1A5A2A03064391. The work of Quoc-Viet Pham was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under Grant NRF-2019R1C1C1006143 and in part by BK21 Four, Korean Southeast Center for the 4th Industrial Revolution Leader Education. Paper no. TII-21-2601. *(Corresponding author: Thippa Reddy Gadekallu.)*

Mamoun Alazab is with the College of Engineering, IT, and Environment, Charles Darwin University, Casuarina, NT 0909, Australia (e-mail: mamoun.alazab@cdu.edu.au).

Swarna Priya RM, Parimala M, Praveen Kumar Reddy Maddikunta, and Thippa Reddy Gadekallu are with the School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India (e-mail: swarnapriya.rm@vit.ac.in; parimala.m@vit.ac.in; praveenkumarreddy@vit.ac.in; thippareddy.g@vit.ac.in).

Quoc-Viet Pham is with the Korean Southeast Center for the 4th Industrial Revolution Leader Education, Pusan National University, Busan 46241, South Korea (e-mail: vietpq@pusan.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3119038.

Digital Object Identifier 10.1109/TII.2021.3119038

to solve the aforementioned limitations of centralized intrusion detection systems as well as other cybersecurity threats.

There have been some surveys over the last few years that focused on providing the fundamentals, applications, and challenges of FL in different research areas. For example, Yang *et al.* [8] presented the fundamentals and application of the FL concept. Lim *et al.* [9] presented a survey of FL at mobile edge networks along with its privacy/security issues and important applications such as computation offloading, edge caching, user association. The importance of FL in IoT and industrial IoT (IIoT) are comprehensively discussed in [10] and [11], respectively. These surveys provide the fundamental knowledge and important lessons for the academic and industrial communities. However, a concise discussion of the application of FL for cybersecurity is still missing. In regard to FL for cybersecurity, there have been some efforts in the literature; however, the existing works focused on specific cybersecurity issues. For example, Mallah *et al.* [12] explored various vulnerable issues in FL-enabled autonomous vehicle systems, and Zhang *et al.* [13] proposed an FL platform and an FL algorithm to efficiently detect the anomaly in IoT systems. To the best of our knowledge, this work is the first attempt to provide a detailed discussion of the use of FL for cybersecurity, from applications and use cases to challenges and scopes for future enhancements.

In this article, we aim to emphasize the importance and applications of FL in dealing with cybersecurity issues along with identifying key challenges and open issues that should be further investigated in the future. The contributions and features of this article can be summarized as follows.

1) *Overview and inspiration*: We start by presenting the fundamentals of FL, cybersecurity, and motivations behind the use of FL for cybersecurity issues. This is detailed in Section II.
2) *Applications of FL in cybersecurity*: Important applications of FL in cybersecurity are provided such as FL for authentication, FL for privacy, FL for trust management, and FL for attack detection. These applications are discussed in Section III.
3) *Potential use cases*: We also present key use cases of FL in cybersecurity, including FL for risk management in finance, FL for objective detection, and FL for antimoney laundering activities. Such potential use cases are presented in Section IV.
4) *Challenges and potential solutions*: We highlight key challenges of FL for efficient cybersecurity systems (e.g., inference attacks, backdoor attacks, and free-riding attacks), and then, discuss potential solutions in Section V.
5) *Future scope and directions*: Finally, we discuss the various open issues and future enhancements like preventing adversarial attacks, building frameworks for FL privacy protection, and hyperparameter tuning in Section VI.

II. State-of-The-Art

Cybersecurity is the major threat nowadays as most of the applications are dependent on the Internet and the data are transmitted from one source to another for taking any kind of sensitive decisions. To build secure intelligent applications, FL could be integrated with sensitive and life critical applications in real time to deal with cybersecurity issues. This section discusses the landscape of FL, cybersecurity, and motivation behind integrating FL with cybersecurity.

A. Landscape of FL

In order to address the limitations of well-trained central ML models like data privacy, confidentiality, inadequate data, high computational power, and high communication cost, the research community tried to develop a framework based on ML called FL [14]. FL successfully overcomes these obstacles by providing a model that trains the data without exposing the actual data. FL is an iterative process in which, during each iteration, the central ML model is improved. The process can be generalized by three major steps as follows.

1) *Selection of model*: A central ML model, called global model, is pretrained with initial parameters, and then, is shared with all the clients in the entire FL environment.
2) *Training locally*: The global model, which is shared along with all parameters to the clients, is trained locally at the client side with their individual data.
3) *Aggregating the local model*: After training locally in the client environment, the updated parameters are forwarded to the central server. Using the updated parameters, the global model is updated. This updated global model is then shared with the clients to start a new iteration.

FL is a framework that learns continuously in an iterative manner by repeating the steps 2 and 3 listed previously and always shares the updated global model to the clients. FL is categorized into horizontal FL (utilized for scenarios where the feature space is same but samples are different) and vertical FL (utilized for scenarios with same sample space but different features). FL is being used in various real-time applications like Gboard, detection of wake word, and analyzing patients' data. In these listed applications, the data are too sensitive, which cannot be shared in a public domain. By utilizing FL, the privacy is maintained locally with the user environment.

B. Cybersecurity

Cybersecurity is the process by which any organization tries to protect their systems that are interconnected and exposed to Internet from various cyberthreats. This practice is followed by individuals as well as enterprises for protecting their data centers and other servers from unauthorized access. Cybersecurity also aims in preventing the external attacks to their sensitive data. A cyberattack is an attack that is made by an individual or organization to breach the information to gain some benefit from disrupting the victim's system. Cyberattacks hit the business every day, and cybercrime has also increased every year. The confidentiality, integrity, and availability (CIA) [15] triad model sets the foundation for all security-based frameworks, which is used for past two decades to enforce security policies. It identifies possible threats and necessary solutions for those threats in the field of information security. The CIA triad model evolves around the following three major components.

1) Confidentiality—Keeping the data secure.
2) Integrity—Keeping the data clean.

Fig. 1. FL for cybersecurity motivations.

3) Availability—Keeping the data accessible.

There are various types of attacks prevailing in cybersecurity that can be categorized based on the intention of attackers and based on the information required for the intruders. Malware is a malicious software, which is activated when the user clicks the unauthorized link or email attachment. Phishing, a common cyberthreat, is the process of sending fraud communications from an authorized source through email to steal the sensitive information. Man-in-the middle is another interesting attack where the intruders will attack in between the network. So, sender will be sending the message without knowing that the hacker in the middle is receiving the information. One of the popular attack DoS, floods the system and the server with traffic to exploit the available resources and bandwidth. Unfortunately, even the legitimate request also cannot be addressed due to DoS attacks. Structured query language (SQL) attack inserts the malicious code into the server and forces the server to reveal sensitive information from the database. A zero-day attack is an unknown security threat in the software application for which the patch has not been released or software developers are unaware of the attack.

C. Motivations of FL for Cybersecurity

Privacy-intrusive way of collecting and sharing data is more challenging in the traditional approach of cybersecurity. Similarly, data consolidation from different data parties is also a complex task. In order to minimize cyberattacks and to achieve data privacy and security, FL could be utilized. Fig. 1 depicts the different factors that influence the use of FL for cybersecurity and methods used in FL to achieve these benefits. The motivations behind utilizing FL for cybersecurity are listed as follows.

1) *Fragmentation and data privacy*: The information of the users is vertically fragmented over different objects in the feature space, where each object keeps track of a specific data feature related to all the users. Each entity in the core network transfers the parameters of the local model trained by locally collected data features rather than sending the raw data to the server, which helps in maintaining the privacy.
2) *Data security*: Securing information from various cyberattacks such as SQL injection, man-in-the-middle and DoS attack can be achieved by utilizing FL since the raw data and information is not communicated over the network, instead only updates are forwarded to the server.
3) *Confidentiality*: Data misuse by any unauthorized access leads to data breach and cyberthreat. Only the authorized people can gain access on the privileged and confidential information. Local training of models in edge devices ensure the authorized access while FL is utilized.
4) *Availability*: Providing access to user information when required must be ensured. Availability is interconnected with reliability and system uptime, which is impacted by malicious issues like cyberattacks and threats. If FL is utilized, the local model is available in the edge device and global model is available in the cloud for user access.
5) *Integrity*: Maintaining consistency, accuracy, and completeness of data is vital part in cybersecurity. Hacker may modify the sender data before it reaches the receiver. In case of FL model, the data are secured as it is designed for privacy preservation and the sensitive data are not transferred out of the local environment.

III. Applications of FL in Cybersecurity

In this section, we discuss how FL can be used for authentication, privacy preservation, intrusion detection, and anomaly detection. Also, some of the recent state-of-the-art on aforementioned works are discussed briefly.

A. FL for Authentication

User authentication (UA) is a decision task that involves accepting or rejecting test inputs based on input training. The similarity is frequently verified in an embedding space, which means that if the expected embedding of the input matches the source embedding, the input is accepted; or else, the input is denied. Authentication models must be trained on a wide variety of data sources in order to understand the various characteristics of data and accurately deny fakers. Collecting user data and training the model in a centralized manner is a difficult task due to security and privacy concerns. Data privacy is essential in UA applications because the model is trained and tested in different environments. The leakage of embedding information decreases the effectiveness of the authentication model by introducing various attacks, such as evasion and poisoning attacks. FL trains ML models with various user data by interacting model weights and patterns between a server and a group of users. In FL, models are trained without sharing users' data with the server or other users. However, training UA models in a federated environment poses challenges.

Developing a reliable FL framework to guarantee authentication has not been extensively researched and remains unresolved. Federated UA (FedUA) was proposed by the authors in [16] for training UA processes. To provide authentication for embedding vectors and inputs, the proposed model employs random binary embeddings and FL. The FL is used to train UA models on

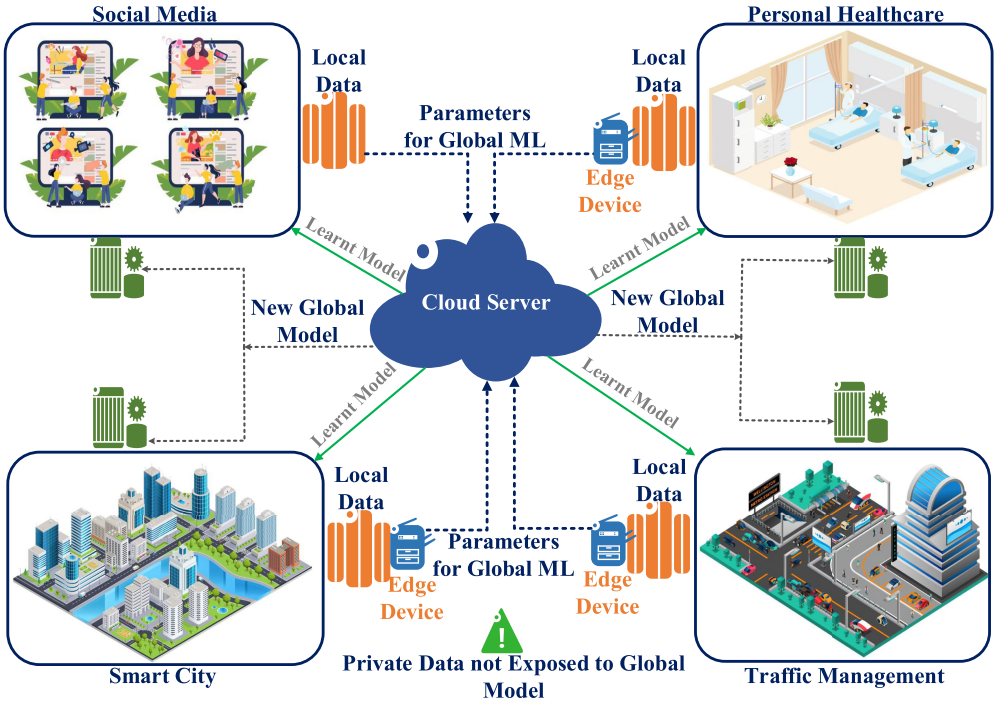

Fig. 2. Privacy preservation through FL.

a massive amount of user's data without direct access to the input data. The authors were successful in achieving scalability with the number of users and do not require any interoperability between the users or among the users and the server, besides the communications made in the FL configuration. The VoxCeleb dataset was used in the experimentation for speaker recognition. This test was carried out by training the speech data of 1, 251 multiple users with various ages, genders, and so on, to accept or decline fakers with high precision and assessing the authentication efficiency on the users' speech data. The experimental results showed that the model trained with FL attained a true positive rate of 80% and a false positive rate of 20%, which was superior when compared to other state-of-the-art models. In [17], the authors proposed VFChain, a blockchain-based FL methodology that provides verifiability and auditability. This procedure consists of several steps. Initially, a blockchain collects models and verifies valid proofs to provide verifiability. In the next step, the authors implemented an authenticated data structure for the blockchain to provide auditability and increase the search efficiency for verified proofs. Following that, an integrative audit layer is used to aggregate all audit files, increasing the system's optimization. The experiments were carried out on the MNIST dataset, which contains 60 000 training samples and 10 000 testing samples. To train the model, the convolution neural network (CNN) is used, with the learning rate requirements set to 0.01 and the batch size set to 100. The experimental results showed that the proposed VFChain outperformed other state-of-the-art models.

B. FL for Privacy

In the traditional ML, the data from individual devices/locations are transferred to the centralized cloud for training the model. This may lead to exposing the sensitive data like patients' medical data, individual's location data, etc., to the potential intruders/hackers. FL enables customized predictions for the individual users by running the predictive algorithms in the local devices, and sends only the parameters that are required to run the global ML model for predictions at larger scale.

Fig. 2 depicts how FL helps in attaining privacy preservation in several applications. In several applications such as social media, smart city applications, healthcare applications, traffic management, etc. sensitive data are stored locally (mobile devices or edge devices). Only the parameters from these devices are sent to the global FL to train the global ML model. Some of the recent works on FL for privacy preservation are discussed as follows.

DL-based approaches have been successfully applied recently on huge volumes of traffic data generated by several organizations to predict the traffic flow. These datasets have sensitive information related to the users. Using these datasets for predictions without exposing the sensitive data is a challenge. To address this issue of privacy preservation of the users' data in traffic flow prediction, Liu *et al.* [18] presented an FL-based gated recurrent unit neural network algorithm (FedGRU). In contrast to the centralized learning approaches, the universal model is updated by the FedGRU by secure aggregated parameters, instead of sharing the raw data from the organizations. To reduce the communication overhead, the authors proposed an enhanced FedAVG method integrated with the joint-announcement protocol for aggregating the parameters from different organizations. The proposed model attained results comparable with the traditional gated recurrent unit (GRU), which is a centralized approach, with improved the privacy preservation. In a similar work, Zhao *et al.* [19] proposed integration of local differential privacy

and FL to address the privacy preservation and communication overhead issues in Internet of vehicles.

Even though FL is a promising solution for privacy preservation in IoT-based applications, several challenges are to be addressed for realizing the full potential of FL for privacy preservation. For instance, the intermediate states of the models can be observed by the clients, and there is a possibility that the clients may arbitrarily update the data in the decentralized process of training the model. Malicious clients may use this opportunity to manipulate the training process with limited or no restriction. These malicious clients may incur model poisoning, in which the malicious clients acting as honest clients can influence the training model's performance by sending erroneous updates [20].

C. FL for Trust Management

The tremendous growth of IoT, smart phones, generates massive amounts of data. ML techniques are used to train the model for providing recommendations and predictions to deliver better mobile services. However, to train the model using ML techniques, a huge amount of user data containing sensitive privacy information must be stored in a server. This increases communication costs, storage space requirements, and the risk of sensitive data privacy leakage and poor trust management. Although FL has many advantages, it is still vulnerable to a variety of security threats in its early stages and lacks expected trust mechanisms. Specifically, data owners may intentionally or unintentionally mislead a global model during an FL process. Sometimes, data owners intentionally send suspicious updates to the global model, causing the current collaborative learning to fail. Thus, it is essential to build effective methods for detecting unreliable local model updates for the FL.

In [21], the authors proposed reliable FL in mobile networks. This process has several steps; initially, a reputation-based worker selection scheme is introduced for selecting a reliable worker. In next step, the authors used a multiweight subjective logic model for a reliable and scalable reputation management. In the final step, to ensure trustable FL, the authors developed a contract theory-based incentive mechanism to enable high-reputation workers with more accurate and trustworthy local training data. The experimental results showed that the proposed contract-based incentive framework produced high-reputation workers with higher quality local training data, ensuring trustworthy FL.

In [22], the authors proposed a privacy-preserving decentralized model ensuring participants' trust in FL. Trusted FL is a decentralized peer-to-peer infrastructure that employs decentralized identifiers and verifiable credentials to accomplish mutual authentication. The Hyperledger Aries modes of communication is used for library deployment and analysis. The experimental results show that the proposed model provides a better trust mechanism than other models.

D. FL for Attack Detection

With rapid increase in digitization and online transactions, sensitive data of the individuals as well as the organizations are under constant threat from potential hackers and intruders. The surge in IoT-based applications has also increased the possibility of sensitive data collected from the sensors being exposed to the malicious users. In order to safeguard the data in a network, a strong intrusion detection system (IDS) is the need of the hour. Anomaly detection is another technique that can be adapted to identify the anomalies in the transactions or the request of data from the networks. ML-based models have been adopted efficiently in the recent past by many researchers to identify the patterns of the intruders, which can help in identifying a malicious user/intruder in the network to build a strong IDS as well as in anomaly detection. If the signatures of the attacks are known, the IDS is a good choice. However, the IDS may fail to detect novel attacks. If the types of attacks are not known, then it is better to use anomaly detection techniques. Anomaly detection mechanisms often suffer from high false detection rates.

In IoT and IIoT-based applications, each node will be equipped with different kinds of sensors. The attacks will be different for every node. The traditional IDS and anomaly detection techniques might not be a viable choice for these applications. The unique feature of FL where the computation of the ML algorithm is executed at the individual devices makes it an ideal solution for detecting the intrusions and anomalies in the aforementioned applications. Through FL, the ML/deep neural network (DNN)-based IDS or anomaly detection algorithms can be executed in the local devices, thereby catering to the needs of providing customized defense mechanisms on those devices. Recent works on application of FL for the IDS and anomaly detection are discussed in the following.

Designing the IDS for dealing with the cyberattacks on heterogeneous, complex, and large-scale cyber-physical systems (CPS) is extremely challenging as the attack examples are significantly less. To tackle this problem, Li *et al.* [4] proposed a novel DeepFed algorithm, which is based on the federated DL. A novel IDS mechanism is designed by the authors based on the CNN integrated with a GRU. A framework based on FL is developed for building a collective IDS from multiple industrial CPSs. For privacy preservation and security of the training process, the authors employed a Paillier cryptosystem-based secure communication protocol. The experimental results on the real-time industrial CPS dataset proved the effectiveness of the proposed model.

Chen *et al.* [23] proposed an IDS mechanism named FedA-GRU, an FL-based GRU, for designing an IDS system for securing wireless edge networks. Instead of sharing the raw data from the edge devices to the central server, FedAGRU updates the universal learning model with the parameters from the edge devices. To increase the weight of the important devices, the authors used the attention mechanism, through which, unimportant updates are not uploaded to the server. In this way, the communication overhead can be reduced and the learning convergence is ensured by FedAGRU. The experimental results prove that the detection accuracy of FedAGRU is improved by 8% compared to the traditional centralized algorithms. Also, the communication cost is reduced by 70% when compared to the conventional FL algorithms.

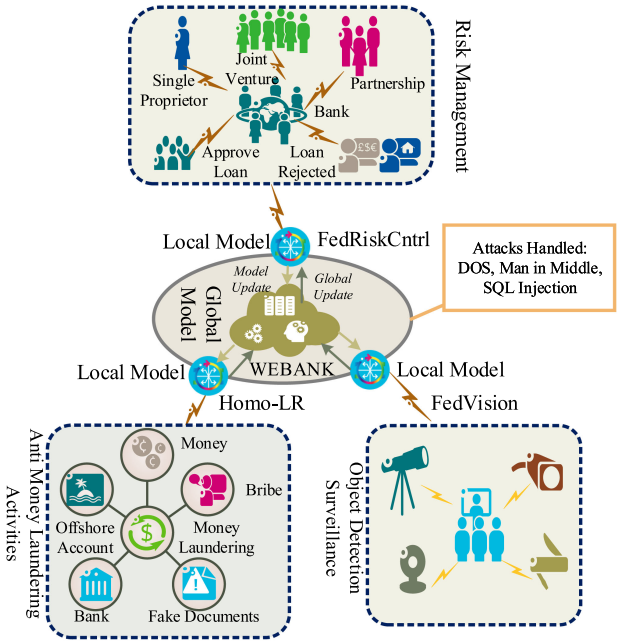

Fig. 3. Illustration of FL-based cybersecurity use cases in WEBANK.

Liu *et al.* [24] proposed an FL-based DNN model for sensing anomalies in time-series data in IIoT. To train the anomaly detection model collaboratively by the decentralized edge devices, the authors used a framework based on FL. Later, they proposed a CNN long short-term memory (LSTM) based on the attention mechanism for detecting the anomalies accurately. Attention mechanism is used in this article to prevent gradient dispersion problem and memory loss in the CNN by capturing important features from the dataset. The LSTM helps in accurate predictions of the time-series data. To improve the communication accuracy, the authors propose a gradient compression mechanism. The results obtained prove that the proposed methodology detects the anomalies in an accurate and timely manner when compared to the traditional algorithms and also the communication overhead is reduced by 50% when compared to the conventional FL model.

IV. Use Cases

FL could be used for building very high end intelligent cross domain, cross data, and cross enterprise real-time applications by allowing the users to maintain their data privately. The FL models have high end utilization in various domains like education, finance, smart city, insurance, edge-computing-based applications and devices, healthcare, and many other similar smart intelligent applications. This section discusses few use cases deployed in real time using the FL technology as shown in Fig. 3.

A. FL in Finance for Risk Management

At present, most of the banks are using a mechanism called whitelisting for managing the risk in small and micro enterprise (SME) loans. The whitelisting is a mechanism that uses some rules for screening purpose using manual intervention. For this purpose, the data have to be retrieved from a central bank that has all the reports with respect to credit. Currently, they use encryption using the Rivest, Shamir, Adleman (RSA) encryption algorithm while transferring the sensitive information. This methodology is used only between authorized agents and banks. If artificial intelligence (AI) has to be built for reducing the manual intervention, ML models need to be used. But gathering initial data based on which the ML model is designed and also for testing is one of the major challenges due to sensitivity of data. Initially, a bank in China named WeBank tried to solve this manual intervention by implementing an AI-based model for risk control. But, transferring sensitive data over Internet was a major challenge due to cyberattacks. To overcome this challenge, they implemented FedRiskCtrl, which is an FL-based risk control mechanism for the application of SME loans. The application is implemented using the Federated Artificial Intelligence Technology Enabler platform. The application is built using the concept of heterogeneous FL. WeBank maintains the model named Hetero-LR and all the other smaller agencies and partners use their own data privately in their location to train the global model Hetero-LR. After training in local environment, the updated parameters will be sent to the centralized WeBank. By implementing this model, WeBank has improved the screening process of SME loans by removing manual intervention and providing security from cyberattacks. This deployment can be clearly understood using a simple instance. Let us consider that Webank has a feature Y and another bank "A" has a feature X1, which is related to risk management. Now, we consider that bank A wants to predict the credit score using all the possible features, including Y. In this case, there are two problems, one is Webank's Y feature is not known to bank A and bank A's X1 feature is not known to Webank. To predict the credit score, either bank needs to transfer the sensitive data over the network, which is insecure in the case of conventional approaches. When Hetero-LR is used, no data are communicated and both banks can train the global model with their own local sensitive features in their own secure environment, thus proving data and model security. When compared to a conventional ML approach, Hetero-LR provides about a 12% increase in prediction accuracy with respect to the screening process.

B. FL in Edge Computing for Object Detection

The surveillance companies usually develop a centralized cloud storage where they store their surveillance videos. But due to privacy concerns and the cost required for transmitting these huge videos to the centralized storage, object detection models developed for detecting the objects fail since they cannot use the videos for training the model, and hence, prediction becomes a challenge in real time. Recently, in China, an online visual object detection platform named FedVision is deployed by association of WeBank with extreme vision. The deployment is for the purpose of fire detection in their environment using the FL concept. The platform works in three major steps namely crowdsourced image annotation, federated model training and federated model update. The crowdsourced image annotation step helps the surveillance company user to locally tag their video in their local machine. This is used for training the global

model retrieved from the edge device instead of transmitting the video to the cloud storage. This step plays a major role in maintaining the privacy of the video and also reduces the cost of transmitting the video over the Internet. The application deployed helps the surveillance company to overcome the risk of any cyberattacks while transmitting the video or while storing in central server. The application uses the horizontal FL concept since the data format is going to be the same but the source of data is different. Due to the usage of horizontal FL, the same model can be used by multiple parties who own same type of data. The FedVision platform employs an approach named YOLOv3 for detecting the objects in a video stream. First, the video stream is partitioned into image frames. Then, each image is divided into *n*n* grids. Each grid is then processed for detecting the target. The major steps involved are prediction of the position of each grid in the entire image, calculation of the confidence score, and finally, conditional probability of the class. The whole process of FedVision is based on the sample-based FL as multiple parties are involved, and every party is sharing a common feature space.

C. FL for Antimoney Laundering Activities

Antimoney laundering activities are usually considered as a vital activity in banking transactions. Traditionally, if a transaction has to be determined as a money laundering activity, the banks use rule-based models for filtering the right records, and then, manual intervention is required for reviewing the transaction record and stamping the record as money-laundering activity, which is time consuming. Though various ML algorithms were used in most of the banks for determining these types of transactions on a daily basis, the performance was not as expected due to lack of huge data for the training purpose. To obtain huge data, multiple banks need to be integrated, but this is not practical as the banks hesitate to share their private data. Even if they share, the attackers might misuse the data so that the activity is misinterpreted. Hence, an online bank in China developed a platform using FL for integrating all the banks virtually and created a model named Homo-LR for training purpose and to prevent from cyberattacks. The banks deal with homogeneous data, that is, the data from multiple banks have the same features but different identities. Hence, by integrating multiple banks, various scenarios of positive cases were evident, which helped develop a global model with all the parameters. Using the global model, every individual bank trained the model locally without sharing the data publicly and the global model was updated with the newer parameters with the help of a third-party aggregator named Arbiter. This concept helped solve the data island problem without sacrificing the data and institutional privacy. Due to this application, the number of transactions needed to be reviewed manually for stamping as money laundering activity reduced from thousand records to 38 records. The percentage of money laundering transactions reduced drastically and the transaction records were secured from attackers.

V. Challenges and Possible Solutions

FL is designed having in mind privacy as a default feature. This is achieved by reducing the footprint of the users' sensitive private data over the network. That is, the users' sensitive data isare not transmitted to a centralized storage. Though FL is claimed to be a privacy-aware ML framework by the researchers, it does not act as an immune system to vulnerable attacks and the current developments are also not much promising to solve all the known privacy issues in a default manner. Motivated by this situation, we explore the existing challenges, current techniques designed for overcoming the issues and also provide an insight on scope for future directions in this section.

A. Inference Attacks

FL is designed to guarantee the privacy of the users and participants. This is achieved by sharing the outcome of the local training done in the user's environment as model parameters instead of sharing their actual sensitive data. But there is still a chance for partially revealing the users' actual training data using the model parameters, which are updated on the centralized global model. These types of threats can be one among membership inference attacks, unintentional data leakage by reconstruction through the inference models, and generative adversarial network-based inference attacks.

Solution: In [25], the privacy of the FL model is improved by integrating the concept of differential privacy (DP) with the traditional shuffling technique and they try to mask the users' data using an algorithm named the invisibility cloak algorithm. But this methodology might reduce the training performance of the FL model as there might be uncertainty in the parameters that are uploaded. VerifyNet [26] is a framework developed for preserving the privacy by providing a double-mask protocol due to which attackers are prevented from predicting or inferring the train data. The researchers need to focus on developing frameworks for preserving the privacy by reducing the communication overhead.

B. Backdoor Attacks

Backdoor attack is a type of attack by which the attackers try to inject a task that is malicious within the existing FL model without affecting the rate of accuracy when the actual task is performed. The process of detecting such attacks is a highly time-consuming process as there is no impact on the accuracy of the model. The backdoor attacks are severe when compared to other attacks as the time taken to detect the occurrence of attack is too high and by the time the attack is detected, the attack would have collapsed the FL model and the false positives are predicted confidently.

Solution: In [27], a methodology named SAFELearning is proposed for enabling the FL model users to detect the backdoor attacks while the aggregation of the model parameters happen. The methodology achieves the aim to detect the backdoor attack by utilizing two techniques named oblivious random grouping and partial parameter disclosure.

C. Adversarial Attacks

In the FL landscape, certain clients are termed as adversarial because they might use their old local data just for the sake of updating the global model once. After the global model is

obtained, these adversarial clients try to deduce the information of other clients in the landscape. This behavior of certain clients are termed as adversarial attacks. These types of behaviors cannot be identified as attacks since the profile regarding clients are limited and are not reputed.

Solution: Two attack models are proposed in [28] named PoisonGAN and DataGen for regenerating the victim's samples using the global model parameters iteratively. These models were experimented on an FL prototype and the results demonstrated that apart from adversarial attacks, the models were also effective on label flipping and backdoor attacks.

D. Free-Riding Attacks

In FL environment, out of all the clients participating in the training process, few clients might act as passive clients. The passive clients do not take part in the training process, and hence, do not contribute. But they will always be connected to the global model and will share the benefits. In some situations, the passive client might try to inject few dummy parameters to update the global model without performing training with their local data. This is termed as free-riding attack. When the FL landscape is smaller, the impact of this type of attack will be severe.

Solution: In [29], a framework named BytoChain is proposed for improving the verification of the models. The verification is done in a parallel manner by introducing verifiers who would verify the models extensively. They use a concept called Proof of Accuracy. The framework is built using blockchain-based FL.

VI. Scope for Future Enhancements

Based on the extensive survey done, this section discusses various challenges still open in the FL domain, which need to be focused on by researchers in the near future for facilitating the adoption of FL models in real time. This can pave the way for researchers on developing a privacy-concerned FL model, which is less prone to cyberattacks.

A. Efficient Backward Traceability

The major challenge in FL privacy and security is the process by which the global ML model could be traced throughout the ML life cycle. The FL global model gets updated by the training parameters of the client after training locally with their own sensitive private data. When these model parameters are updated, there is a chance that the prediction value is completely changed in the global model. When such change occurs, there should be some mechanism using which the model can backtrack the process so that the process can identify the client whose aggregated model parameter values caused the change in prediction value of the global model. If such a transparent backtracking model is unavailable, then the ML model logic would become a black box and the users will be forced to lose the logical reality. This would result in going behind the unreality of the human- made AI decisions and predictions. The future researchers can focus on building or developing a more transparent traceable training model for ML process.

B. Developing Process and Standards

FL is a newly blooming technology and it requires a high level of descriptive analysis of the various advantages and disadvantages with respect to various approaches. The researchers need to focus on developing standard techniques for supporting the upcoming requirements of the FL in various domains. Since privacy is a major motivating factor, focus has to be toward improving the privacy and developing standard approach for every requirement. A process has to be defined supporting the generic application programming interface (APIs) for implementation of the enhancements. The researchers need to focus on developing new standards for implementing the preventive mechanisms for cyberattacks. This can enhance the trust among the clients and the global model providers.

C. Building Frameworks for FL Privacy Protection

Currently there are very few frameworks like TensorFlow Federated, PySyft, and FATE for the purpose of implementing FL-based applications in real time. Out of these frameworks, only PySyft is designed to integrate the DP concept for providing privacy preservation. Hence, the researchers can focus on developing privacy preservation frameworks, which can be advantageous for both academicians and industrialists in adopting the FL for cybersecurity.

D. Techniques for Hyperparameter Tuning

The research community needs to concentrate on developing optimization techniques for deciding the start and end criterion for the training phase. Also proposing mechanisms for tuning the hyperparameters while configuring the training process is a promising thrust area. Moreover from the study, we have understood that the FL training process is too time consuming and the computational cost is also a bit higher than the traditional ML techniques. Development of cost-effective and time-efficient FL models can be very supportive.

E. Handling 5G and Beyond Networks

5G and beyond networks would help in providing connectivity to almost all types of devices, networks, places, and things related to daily activities [30]. For instance, 5G vision is toward interconnection of almost all physical entities in the real world like schools, industries, automobiles, marketing, retail, banks, hospitals, cities, etc. These interconnected networks are prone to various security threats like DoS attacks, hijacking attacks, signalling storms, resource theft, configuration attacks, saturation attacks, penetration attacks, user identity theft, transmission control protocol (TCP) level attacks, man-in-the middle attack, reset and IP spoofing, scanning attacks, insider attacks, data leakage, cloud intrusion, active eavesdropping, and passive eavesdropping. The threats are not restricted to these listed ones. When 5G and beyond networks are commercially implemented, there are many more unknown vulnerabilities. Utilization of FL for cybersecurity supports the designers in solving certain threats. The future focus has to be in developing mechanisms for handling all the security threats and vulnerabilities.

F. Other Future Enhancements

The transformative evolution paved a path to Industry 5.0, which integrates collaborative robotics and human into a hyperconnected system. Industry 5.0 is characterized by automation and extreme exchange of information throughout the business chain. Cybersecurity risks are vital in Industry 5.0 due to the high cyberthreats caused during the exchange of information. Risk mitigation strategies and privacy mechanisms can lower the risk of cyberthreats. The vulnerability caused by this interconnection, exchange of information, and advancement in digital transformation can be solved by implementing FL with cybersecurity concepts. FL exhibits an inherent nature in privacy preservation as data are processed locally. However, the central model aggregation and communication with the clients make FL more vulnerable to external attacks. A potential solution to overcome these cyberattacks is integrating the blockchain concept with FL. Blockchain associated with decentralized FL would enhance the security of FL in terms of privacy preservation and tamper resistance data. Another promising solution is combining the power of cybersecurity and crowdsourcing. A crowd can significantly provide high efficiency against cybersecurity threats than a single entity.

VII. Conclusion

FL is a new breed of AI that advocates the on-device AI with the logic of decentralized data learning and extensive training of the prediction models using the sensitive private data of the user. This article provided an extensive survey on how FL could be adopted for cybersecurity. The survey started with the discussion on key concepts in this survey like FL, workflow, cybersecurity, and cyberattacks. We then provided a detailed analysis of the current developments using the FL perspective for providing authentication, privacy, trust management, and attack detection. To motivate and to give a better view and clear understanding of the FL framework in real-time adoption, various use cases deployed in finance, computer vision, and money laundering activities were discussed. To enhance the research in the newly launched FL framework, open challenges were discussed along with the current solutions proposed by the research community. To provide a road map for the future researchers, this article is concluded with future scope and directions.

References

[1] D. Silver *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[2] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtarik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," in *Proc. NIPS Workshop Private Multi-Party Mach. Learn.*, Barcelona, Spain, 2016, pp. 1–10.

[3] Q.-V. Pham, M. Zeng, R. Ruby, T. Huynh-The, and W.-J. Hwang, "UAV communications for sustainable federated learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3944–3948, Apr. 2021.

[4] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated deep learning for intrusion detection in industrial cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5615–5624, Aug. 2021.

[5] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "FedHealth: A federated transfer learning framework for wearable healthcare," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 83–93, Jul./Aug. 2020.

[6] S. R. Pokhrel and J. Choi, "Federated learning with blockchain for autonomous vehicles: Analysis and design challenges," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4734–4746, Aug. 2020.

[7] S. A. Rahman, H. Tout, C. Talhi, and A. Mourad, "Internet of things intrusion detection: Centralized, on-device, or federated learning," *IEEE Netw.*, vol. 34, no. 6, pp. 310–317, Nov./Dec. 2020.

[8] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[9] W. Y. B. Lim *et al.*, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surv. Tut.*, vol. 22, no. 3, pp. 2031–2063, Jul.–Sep. 2020.

[10] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 1622–1658, Jul.–Sep. 2021.

[11] M. Parimala *et al.*, "Fusion of federated learning and industrial Internet of Things: A survey," 2021, *arXiv:2101.00798*.

[12] R. A. Mallah, G. Badu-Marfo, and B. Farooq, "Cybersecurity threats in connected and automated vehicles based federated learning systems," 2021, *arXiv:2102.13256*.

[13] T. Zhang, C. He, T. Ma, M. Ma, and S. Avestimehr, "Federated learning for Internet of Things: A federated learning framework for on-device anomaly data detection," 2021, *arXiv:2106.07976*.

[14] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[15] P. Zhuang, T. Zamir, and H. Liang, "Blockchain for cybersecurity in smart grid: A comprehensive survey," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 3–19, Jan. 2021.

[16] H. Hosseini, S. Yun, H. Park, C. Louizos, J. Soriaga, and M. Welling, "Federated learning of user authentication models," 2020, *arXiv:2007.04618*.

[17] Z. Peng *et al.*, "VFChain: Enabling verifiable and auditable federated learning via blockchain systems," *IEEE Trans. Netw. Sci. Eng.*, to be published, doi: 10.1109/TNSE.2021.3050781.

[18] Y. Liu, J. James, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7751–7763, Aug. 2020.

[19] Y. Zhao *et al.*, "Local differential privacy based federated learning for Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 8836–8853, Jul. 2021.

[20] C. Ma *et al.*, "On safeguarding privacy and security in the framework of federated learning," *IEEE Netw.*, vol. 34, no. 4, pp. 242–248, Jul./Aug. 2020.

[21] J. Kang, Z. Xiong, D. Niyato, S. Xie, and J. Zhang, "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10700–10714, Dec. 2019.

[22] P. Papadopoulos, W. Abramson, A. J. Hall, N. Pitropakis, and W. J. Buchanan, "Privacy and trust redefined in federated machine learning," *Mach. Learn. Knowl. Extraction*, vol. 3, no. 2, pp. 333–356, 2021.

[23] Z. Chen, N. Lv, P. Liu, Y. Fang, K. Chen, and W. Pan, "Intrusion detection for wireless edge networks based on federated learning," *IEEE Access*, vol. 8, pp. 217463–217472, 2020, doi: 10.1109/ACCESS.2020.3041793.

[24] Y. Liu *et al.*, "Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6348–6358, Apr. 2021.

[25] B. Ghazi, R. Pagh, and A. Velingker, "Scalable and differentially private distributed aggregation in the shuffled model," 2019, *arXiv:1906.08320*.

[26] G. Xu, H. Li, S. Liu, K. Yang, and X. Lin, "VerifyNet: Secure and verifiable federated learning," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 911–926, 2020, doi: 10.1109/TIFS.2019.2929409.

[27] Z. Zhang, J. Li, S. Yu, and C. Makaya, "SAFELearning: Enable backdoor detectability in federated learning with secure aggregation," 2021, *arXiv:2102.02402*.

[28] J. Zhang, B. Chen, X. Cheng, H. T. T. Binh, and S. Yu, "PoisonGAN: Generative poisoning attacks against federated learning in edge computing systems," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3310–3322, Mar. 2021.

[29] Z. Li *et al.*, "Byzantine resistant secure blockchained federated learning at the edge," *IEEE Netw.*, vol. 35, no. 4, pp. 295–301, Jul./Aug. 2021.

[30] C. de Alwis *et al.*, "Survey on 6G frontiers: Trends, applications, requirements, technologies and future research," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 836–886, Apr. 2021, doi: 10.1109/OJCOMS.2021.3071496.