

Securing Federated Learning: A Covert Communication-based Approach

Yuan-Ai Xie, Jiawen Kang, Dusit Niyato, *Fellow, IEEE*, Nguyen Thi Thanh Van, Nguyen Cong Luong*, Zhixin Liu, and Han Yu

***Abstract*—Federated Learning Networks (FLNs) have been envisaged as a promising paradigm to collaboratively train models among mobile devices without exposing their local privacy data. Due to the need for frequent model updates via wireless links, FLNs are vulnerable to various attacks (e.g., eavesdropping attacks, replay attacks, inference attacks, and jamming attacks). Balancing privacy protection with efficient distributed model training is a key challenge for FLNs. Existing countermeasures incur high computation costs and are only designed for specific attacks on FLNs. In this paper, we bridge this gap by proposing the Covert Communication-based Federated Learning (CCFL) approach. Based on the emerging communication security technique of covert communication which hides the existence of wireless communication activities, CCFL can degrade attackers' capability of extracting useful information from the FLN training protocol, which is a fundamental step for most existing attacks, and thereby holistically enhances the privacy of FLNs. We experimentally evaluate CCFL extensively under real-world settings in which the FL latency is optimized under given security requirements. Numerical results demonstrate the significant effectiveness of the proposed approach in terms of both training efficiency and communication security.**

***Index Terms*—Federated learning, secure aggregation, covert communication, privacy attacks.**

I. Introduction

Federated Learning Networks (FLNs) [1] have been proposed as a distributed privacy-preserving collaborative model training approach to alleviating societies' concerns on the exposure of sensitive data when building artificial intelligence applications. In FLNs, a server and a large number of mobile devices (MDs) perform multiple rounds of training iterations through wireless model updates to build machine learning models for specific tasks [2].

However, to implement FL over wireless networks, the MDs must transmit their local training results over wireless links [3] that are vulnerable to various types of wireless attacks. These wireless attacks include the eavesdropping attacks, replay attacks, inference attacks, and jamming attacks, which in turn, have inspired a myriad of defense mechanisms (a.k.a. countermeasures) [4]. The existing countermeasures have some limitations. Firstly, given that each countermeasure is designed to address a specific attack, it is costly to deploy multiple countermeasures to tackle multiple types of potential attacks on FLNs. Secondly, the countermeasures might interfere with each other, which can further complicate FLN security issues. For example, a countermeasure to jamming attacks requires the higher transmit power that leads to more successful eavesdropping attacks. Hence, it is necessary to propose a unified, efficient, and highly secure solution to combat a broad range of attacks in FLNs. We observe that attacks are always launched under a common premise that the attacker is aware of the existence of the transmission involved in training FLNs. Yet, this fundamental premise has not been leveraged by existing FLN defense mechanism research.

Recently, Covert Communication (CC) has been introduced as a promising security technique to prevent adversaries from detecting the existence of wireless transmission links [5], [6], [7]. Historically, the spread spectrum technique was adopted to achieve CC through spreading the transmitted signal power over a large time-frequency space. However, its covertness cannot be well analyzed. Hence, channel artifacts, such as additive white Gaussian noise channels, are used to hide communications. The fundamental information-theoretic limits of CC over random channels (i.e., the square root law) were explored in [8]. Specific CC techniques, such as Artificial Noises (AN) or jamming signals, were widely used to prevent attackers from detecting the legitimate transmissions [5], [7]. Compared with the traditional cryptography and Physical Layer Security (PLS) technologies, CC can provide higher-level security by hiding transmissions that attract attackers' attention [6].

As perceiving whether a wireless transmission exists or not is often the first step for external malicious third parties to launch attacks on FL, we envision a *Covert Communication-based Federated Learning (CCFL)* approach to secure the model transmissions from the MDs to the FL server as a cost-effective fundamental approach for precluding such attacks. When re-contexting CC into FLNs, key technical challenges such FL latency increase and synchronization issues for the combination need to be resolved. Hence, CCFL must jointly optimize the transmit power of each MD, the jamming power of the friendly jammer, and the local model accuracy at the MDs in a distributed way. This envisioned approach contributes to the FL literature in the following ways:

Yuan-Ai Xie and Zhixin Liu are with the Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China. Email: xieyuan_ai@163.com, lzxauto@ysu.edu.cn.

Jiawen Kang is with the School of Automation, Guangdong University of Technology. Email:kavinkang@gdut.edu.cn.

Dusit Niyato, and Han Yu are with School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798. Emails: dniyato@ntu.edu.sg, han.yu@ntu.edu.sg.

Nguyen Thi Thanh Van is with the Faculty of Electrical and Electronic Engineering, PHENIKAA University, Hanoi, Vietnam. Email: van.nguyenthithanh@phenikaa-uni.edu.vn.

Nguyen Cong Luong* (Corresponding author) is with the Faculty of Computer Science, PHENIKAA University, Hanoi, Vietnam. Email: luong.nguyencong@phenikaa-uni.edu.vn.

1) **Holistic Security:** Through exploiting the common premise for launching various attacks, CCFL provides a unified and holistic security framework to mitigate a broad range of attacks on FLNs.
2) **Cost Effectiveness:** As CCFL is aimed at the key enabling step for malicious third parties to launch attacks on FLNs, it can preclude such attacks. In this way, MDs are no longer required to host computationally expensive countermeasures. This enables more resource-constrained devices to participate in FLNs.

We investigate a case study to show the benefits of CCFL in which the AN-based CC technique is leveraged to hide the model transmissions from the MDs to the FL server from a warden (i.e., an attacker). Numerical results demonstrate significant advantages of the envisioned CCFL approach. The major contributions of this paper are summarized as follows.

- Unlike the traditional point-to-point communication scenarios in which CC was typically deployed, we apply CC to more complex multipoint-to-point FLNs where a massive number of MDs transmit their models to one FL server.
- CCFL can provide a unified, efficient, and highly secure approach to preclude a broad range of attacks on FLNs.
- To enhance the covertness of the wireless model updates, a full-bandwidth jammer is introduced which increases the FL latency. Hence, an FL latency minimization problem is formulated by jointly determining the jamming power, local model transmission power, and local training accuracy under a covertness requirement.

The rest of the paper is organized as follows. In Section II, we introduce common attacks in FLNs. In Section III, we present covert communication techniques and a case study of using the covert communication for FLNs. The conclusions and future works are discussed in Section IV.

II. Overviews of Federated Learning Networks

A. Fundamentals of FLNs

In a typical centralized learning network, the MDs are required to upload their local data to the central server through wireless links. Then, machine learning models are trained in the server (e.g., through Stochastic Gradient Descent (SGD) [1]). Nevertheless, the broadcast nature and limited spectrum of wireless networks as well as centralized data storage have led to critical issues including risks of privacy leakage, high communication overhead, and limited scalability.

To address the issues, FLNs have been proposed. As shown in Figure 1, the MDs obtain a shared global model broadcast by the FL server. They then train the local models with their data, and upload the local model parameters (e.g., gradients) to the server. After that, the server updates its global model by aggregating the received model updates (e.g., through federated averaging [9]). These steps are repeated until the global model converges. During this process, the MDs transmit the model parameters instead of their local data. As a result, FLNs significantly reduce communication overheads and the risk of the privacy leakage.

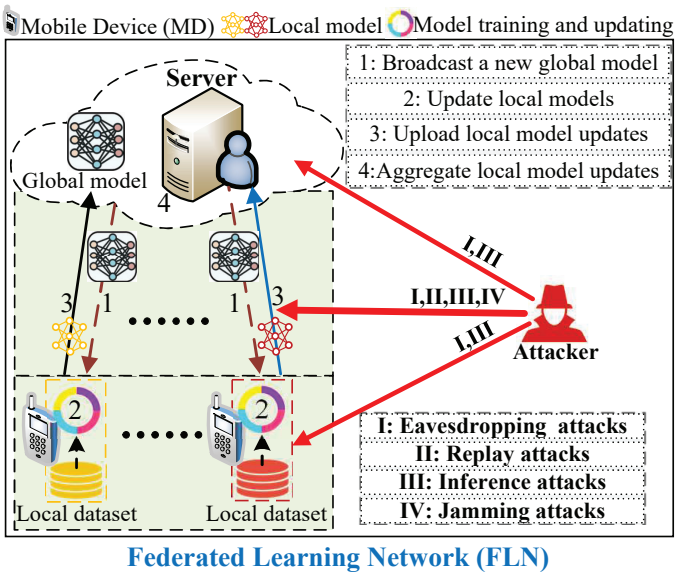


Fig. 1. The architectures of FLNs and their major attacks.

B. Major Attacks on FLNs and Countermeasures

To leverage the mobile edge computing platform and exploit the massive data distributed over a large number of MDs at the edge of wireless networks, wireless communications is used to transport the model updates [2], [3] due to the shared and easy to access wireless medium. However, the exposed nature of the wireless medium makes wireless networks vulnerable to various attacks. The attacks such as eavesdropping attacks, replay attacks, inference attacks, and jamming attacks have been successfully mounted against FLNs [4]. In the following part of this section, we present major attacks in FLNs and discuss the corresponding countermeasures.

Eavesdropping Attacks: An attacker (a.k.a. eavesdropper) eavesdrops legitimate communications between a transmitter and a receiver. In FLNs, the trained models can leak sensitive information about the owners of the MDs (e.g., gender, occupation, and location) [10]. In addition, the attacker can obtain "free" model updates transmitted between the FL server and the MDs, thus allowing them to free-ride without contributions. Since eavesdropping attacks are relatively easy to perform and can escalate to more severe cyber-attacks (e.g., Denial of Service (DoS)[1] and jamming[2]), they are considered as one of the most common and serious attacks.

Countermeasures: To mitigate eavesdropping attacks, cryptographic methods and PLS have been proposed. The cryptographic methods encrypt the transmitted models through a secret key that is only known by its intended receivers (e.g., the FL server). However, these methods also incur high computation costs and system complexity. This is especially challenging for FLNs involving a large number of MDs. Unlike the cryptographic methods, PLS exploits the randomness of wireless channels and the AN (i.e., the jamming signal)

[1]DoS can overwhelm the server by making it go offline and deny further connection requests.

[2]Jamming can cause a poor model accuracy or even interrupt model transmissions.

to limit the number of models extracted or intercepted by an eavesdropper. Nevertheless, PLS cannot provide adequate security since the attackers are still able to capture part of the confidential FL model by side-channel analysis [6].

Replay Attacks: An attacker intercepts data packets transmitted from a transmitter to a legitimate receiver while fraudulently delaying or resending the packets [11]. In FLNs, an attacker can intercept the model updates and replace them with malicious model updates, thereby causing performance degradation through a deceptive delay or re-transmission.

Countermeasures: To prevent such attacks, both the FL server and MDs should establish a completely random authentication code, which is only valid for one transaction [11]. Another countermeasure is using timestamps on all model updates. This prevents attackers from resending FL models that are updated longer than a certain time. This reduces the opportunity for an attacker to eavesdrop, siphon off the models, and resend the modified models. The main drawback of such approaches is the high computation cost, especially for the massive MDs.

Inference Attacks: Inference attacks aim to reveal sensitive information by probing a machine learning model with much hidden information. They fall largely into two categories: 1) membership inference attacks and 2) property inference attacks [4]. The membership inference attacks aim to determine whether an exact data point was used to train a given model. By observing SGD-based gradient updates, attackers can infer a significant amount of private information, and then may launch a powerful attack (e.g., gradient ascent attack) against other MDs. On the other hand, property inference attacks aim to infer properties of training data that are independent of the characterized features of a class. Meanwhile, the attacker is assumed to have auxiliary training data correctly labeled with the property they intend to infer.

Countermeasures: Differential Privacy (DP)-based solutions and encryption-based solutions are commonly used to mitigate inference attacks on FLNs. For DP-based solutions, a rigorous randomization mechanism (e.g., a Gaussian noise mechanism), is designed to inject additive noises into the trained parameters before they are uploaded to the FL server [10]. This guarantees that the addition or removal of a single data sample or model parameter does not affect the outcome of any inference. For example, [12] introduced a differentially private SGD algorithm that can effectively protect the privacy of the parameters. However, due to the noise added to the local models, the overall model accuracy suffers. Encryption-based solutions leverage encryption techniques to secure the data privacy of the MDs when the local model parameters are shared. On this basis, [13] proposed a homomorphic encryption-based technique, which can protect sensitive information while preserving model performance. Nevertheless, they incur high computation costs and require complex system designs.

Jamming Attacks: With jamming attacks, an attacker transmits jamming signals with a regulated power to disrupt the legitimate communications between a transmitter and its receiver [14]. Generally, jamming attacks are launched at the physical layer on a single channel or multiple channels. They can significantly degrade the network performance since they decrease the Signal-to-Interference plus Noise Ratio (SINR) at the receiver. In FLNs, jamming attacks can interfere with the local update transmissions from the MDs to the server. This results in the low SINR at the server, and the server even be unable to decode the signals transmitted from the MDs. Consequently, the global FL model might not be properly updated, and the training performance might degrade.

Countermeasures: The following countermeasures have been proposed to combat jamming attacks on FLNs [14]:

- *Local update transmission with high power:* With high power, the MDs can transmit their local models better and yield a high SINR at the server.
- *Directional antennas:* With directional antennas, the Base Station (BS) at the server can directionally receive signals transmitted from the MDs rather than the signals coming from any direction. This can reduce the interference caused by the jamming signals from the attacker.

Existing countermeasures can protect FLNs against the corresponding attacks to different extents. However, as summarized in Table I, all countermeasures have some limitations when addressing these diverse attacks. Firstly, the MDs are energy-constrained, the anti-jamming methods such as high transmitted power can greatly reduce the lifetime of the MDs, especially for the encryption-based methods with high power consumption. Secondly, the scarce spectrum resources make frequency hopping costly, e.g., a large number of MDs are involved in training FLNs. Thirdly, directional antenna methods can incur high hardware costs at both MDs and the FL server. Hence, a more cost-effective security solution is required. Moreover, in the worst-case scenario where multiple types of attacks are launched simultaneously, the detection of such attacks and deployment of countermeasures might result in unintended complications in addition to the prohibitively high resource requirements and system complexity. Thus, a unified, efficient, and highly secure FLN protection framework is needed for this technology to become widely adopted.

III. Covert Communication-based Federated Learning

In this section, we apply CC to secure FLNs. We first introduce the overviews of CC and some advanced CC techniques. Then, we give a detailed discussion about the superiority of CC and some technical challenges when CC is deployed into FLNs. Finally, we design a CCFL approach for securing FLNs.

A. Covert Communication

CC, a.k.a. low probability of detection communication, aims to mask the existence of a legitimate wireless transmission from a watchful adversary under the requirement of a certain covert rate for the intended user [5], [8]. Generally, CC provides three major advantages. Firstly, different from PLS which prevents an adversary from knowing the messages sent by the transmitter, CC circumvents an adversary from knowing whether the transmission has occurred. If the adversary cannot detect the transmission, it will be unable to launch further

TABLE I
SUMMARY OF VARIOUS HIGH-RISK ATTACKS ON FEDERATED LEARNING NETWORKS (FLNS) AND THE CORRESPONDING COUNTERMEASURES

Attack types	Attack effect	Source of Vulnerability	Countermeasures	Limitations
Eavesdropping attacks	Medium but prevailing	MDs,wireless transmission, compromised server	Cryptographic methods/PLS	High computation costs/ inadequate security
Replay attacks	High	Wireless transmission	Random authentication code/timestamps	High computation costs
Inference attacks	High	MDs,compromised server	Differential privacy-based protection/ encryption-based solutions	Relatively poor model accuracy/ high computation costs
Jamming attacks	High	Wireless transmission	Local update transmission with high power/ directional antennas	Energy-consuming/ high hardware costs

attacks. As a results, CC can combat with several wireless attacks simultaneously and effectively. Secondly, unlike encryption technologies, CC is low-cost and low-complexity since no encryption/decryption algorithms are required. Moreover, its performance does not depend on the adversary's computation capability. Thirdly, CC has wide compatibility and can be easily adopted to complement advanced distributed artificial intelligence techniques, such as FL.

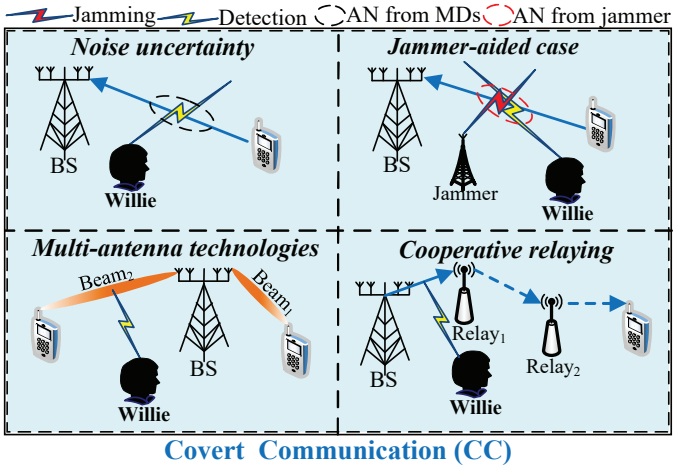


Fig. 2. Major covert communication techniques.

However, CC without extra covertness techniques may not be secure enough, especially when the adversary possesses strong detection capability. To improve the covertness of wireless links, as shown in Figure 2, various approaches have been developed based on the CC techniques [5].

1) *Noise Uncertainty:* Two major sources of uncertain noises are leveraged: background noise and random AN [5]. The background noise is affected by environmental factors such as temperature and humidity, while the random AN can flexibly and efficiently amplify interference dynamics and confuse the adversary [5]. These noises can be uncertain to the attacker, and thus they can be exploited to improve communication link covertness.
2) *Multi-Antenna Technologies:* By adequately exploiting spatial degree of freedom, multi-antenna technologies can help enhance the covertness of wireless links from all directions [15]. Its realization requires beamforming to generate spatial selectivity. Specifically, a beamformer adjusts the corresponding amplitude and phase of the signals on each element of an antenna array in such a way that the superimposed radiation pattern is constructive in the desired direction and destructive in other directions. As a result, the transmitted signals can reach the desired receiver to enhance the data rate and simultaneously null the transmission at the adversary site. As the number of antennas increases, the antenna array will have a higher beamforming resolution which can be utilized to achieve a more reliable covert rate.
3) *Jammer-Aided Technologies:* There can be two types of friendly jammers for enhancing the stealthiness: 1) scheduled jammers, and 2) random jammers [7]. A scheduled jammer is informed about the transmission of the legitimate transmitter and releases its AN with optimized parameters (e.g., jamming power). In contrast, a random jammer is unaware of the transmission by the legitimate transmitter and rather randomly or continuously transmits its AN. Compared with the random jammer, the scheduled jammer is more efficient and has more reliable covertness performance. For example, at the time the legitimate transmitter starts to transmit a codeword, the scheduled jammer turns down the power of the transmitted Gaussian noise, which is turned back up at the moment the transmitter finishes transmitting.
4) *Cooperative Relaying:* By leveraging the cooperation from intermediate node(s), cooperative relaying can achieve CC [5]. Note that the access distance significantly effect the covertness. For long-distance communication, high transmit power is required to achieve a target rate, which unavoidably impairs the covertness. Thus, multi-hop forwarding-based cooperative relaying can be used. The fundamental is to shorten the communication distance of each hop to maintain the required transmit power low, leading to a low detection probability of the adversary. Through this technique, the covertness performance can be considerably enhanced.

Through the comparison in Table I, CC demonstrates its superiority compared with the existing countermeasures. In particular, CC is more efficient and can provide a higher level of security. Moreover, CC is low cost since no encryption/decryption algorithm is required. This is important sicne the MDs in FL are limited in energy and computation that prohibits them from performing complicated encryption algorithms. As a result, CC is a suitable security solution for FLNs. However, there are still challenges needed to be resolved when CC is applied. In particular, the enhancement of system covertness may be at the cost of other performance metrics, such as latency, transmission rate, and global accuracy. For instance, CC techniques mainly exploiting extra noises and jamming signals inevitably cause a lower SINR/a

high training latency and reduce the accuracy of the global FL model. To address the high latency while guaranteeing the CC requirement, we investigate a resource optimization of FLNs and forge a more efficient system. Especially, we consider a case study where the transmit power of each MD, jamming power of the friendly jammer, and local accuracy at the MDs are jointly optimized to minimize the overall FL latency while guaranteeing a security requirement.

B. The Envisioned CCFL Approach

In FLNs, the learning process involves multiple rounds of communications between MDs and the FL server. Adversaries can launch eavesdropping attacks and extract model parameter information via a weak channel condition. By detecting the existence of model updates, it is possible for eavesdropping attacks to escalate into more severe forms of attacks (e.g., DoS, jamming, and replay attacks) to manipulate FL model updates and aggregations.

To curb these attacks effectively and preclude them from escalating, we envision incorporating CC into the FLN training process to hide the occurrence of model update transmissions. Without awareness of these transmissions, it is hard for the adversary to lunch attacks effectively. Since the large-scale MDs are always configured with orthogonal channels, CC can deliver holistic security for the FL MDs through distributedly deploying it to each orthogonal channel. In addition, the distributed power control for CC incurs lower computation costs than existing countermeasures such as cryptographic methods and PLS. Multi-antenna technologies and cooperative relaying are more suitable for the downlink global FL model broadcast. The noise uncertainty and jammer-aided techniques are useful for securing the more vulnerable uplink FL model updates.

C. A Case Study of CCFL

We show a case study in which the jammer-aided technology is used to secure the local models transmitted from the MDs to the FL server.

1) System Model: We consider an FLN as shown in Figure 3 which consists of N devices, one FL server located at a BS, and an attacker Willie. To achieve high efficiency, the orthogonal frequency-division multiple access technique is used for local model uploading by the MDs. A friendly jammer with N antennas is deployed to transmit AN signals continuously with total power p^j to Willie. To secure the model transmissions of all the devices simultaneously, time/clock synchronization algorithms are firstly adopted before the MDs transmit their local models to the BS, and then the jammer leverages the barrage jamming technique (i.e., the jammer can transmit the AN signals over the full bandwidth). Note that the AN signals also cause interference to the BS and may reduce the SINR at the BS. As a result, the data rate of the MDs decreases and the total training latency decreases. The jammer is self-interested and rational. Thus, the server needs to pay the jammer a fee for the jamming service. For simplicity of discussion, a linear cost model is adopted in which the cost paid to the jammer is linearly proportional to p^j.

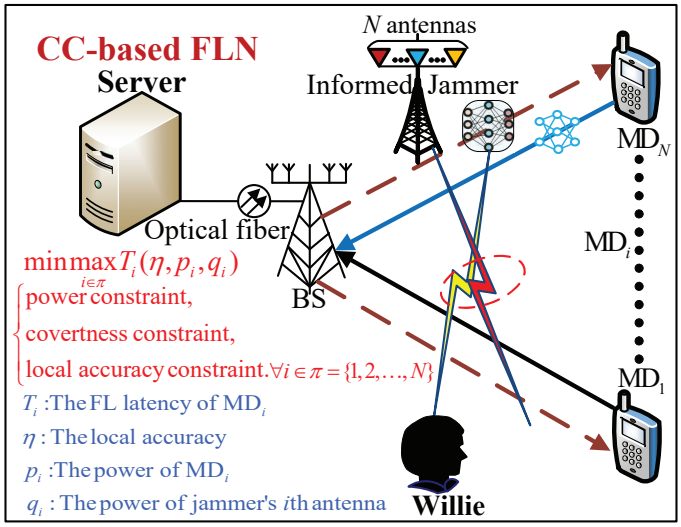
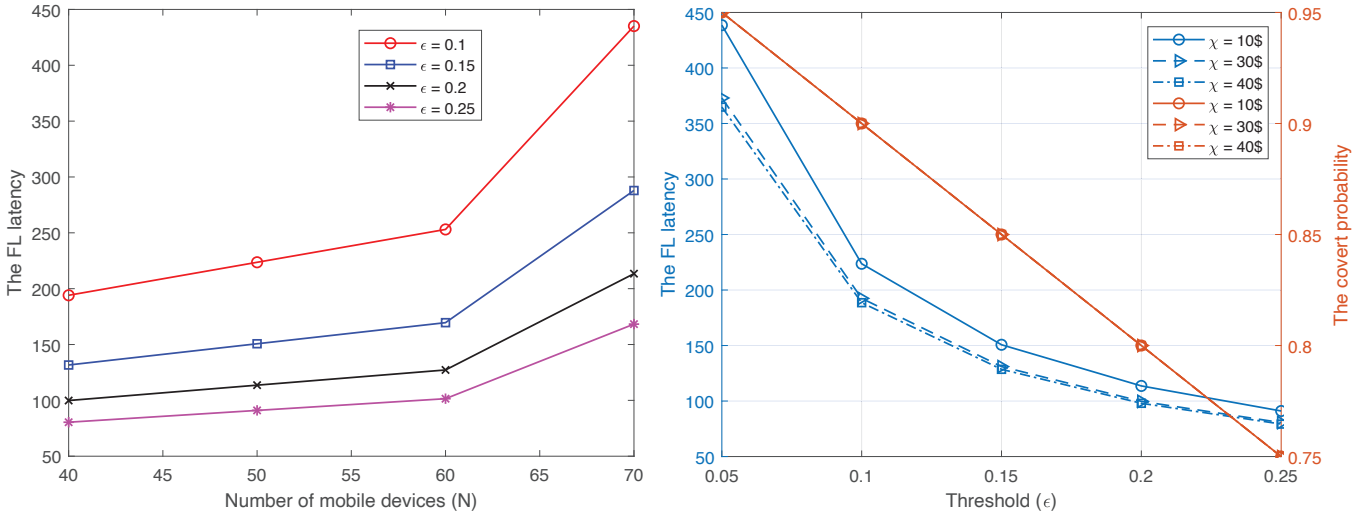


Fig. 3. Covert communication-based model updates for secure FLNs.

The FL model training process involves multiple iterations. In each iteration, the MDs train their local models to achieve a local accuracy η (in terms of training error). At the end of each iteration, each MD can decide to transmit or not to transmit its local model to the FL server with a pre-defined transmission probability. When a device decides to transmit its model update to the server, and Willie judges that the device does not execute the transmission, then a *miss-detection* occurs. When Willie judges that the device is transmitting the update while the device does not, then a *false alarm* occurs. We define the covert probability for a device as the sum of the false alarm probability and the miss detection probability. We expect a high covert probability for situations in which Willie cannot correctly detect the model transmission of any device in the network. For this, the cover probability for the device in the FL network needs to be greater than a security requirement $(1-\epsilon)$, where ϵ is the security threshold. This is the CC constraint.

To prevent Willie from detecting the local model transmissions by the devices, the server can request the jammer to transmit the AN signal with a higher power. However, this increases the cost that the server needs to pay the jammer and also reduces the SINR at the BS, leading to an increase in FL latency. Otherwise, the server can increase the local accuracy η at the devices to reduce the number of local iterations, thereby reducing the computation time at the devices. However, this approach requires more global iterations to achieve high accuracy, thereby increasing the FL latency. Note that how the AN signal impacts on the global accuracy is not considered in this work. Therefore, the joint optimization problem must simultaneously consider: 1) jamming power, 2) transmit power of the devices, and 3) local accuracy at the devices, to minimize the FL latency, subject to: 1) the CC constraint, 2) the maximum power of the devices, 3) the maximum power of the jammer, and 4) the FL server's budget. Here, the FL latency is defined as the maximum latency among the devices. The objective function and the CC constraint are non-convex. Thus, the optimization problem is non-convex.



(a) FL latency versus the number of devices N

(b) FL latency and security performance versus security threshold ϵ

Fig. 4. The number of devices is set as $N = 50$. The devices, jammer, and Willie are distributed randomly in a square area of size of 500 m $\times$ 500 m. The transmission probability of the devices is 0.7, and their maximum power is 10 dBm. Each device has 500 data samples for its local training, and the device has a maximum computation capacity of 2 GHz. The total bandwidth for the model transmissions of the devices is 20 MHz. The price per jamming power unit is 0.5\$ that is set by the jammer, and the budget of the server is 30\$. The security requirement is $\epsilon = 0.1$.

To solve the problem, we can adopt an alternating descent algorithm. The algorithm divides the original problem into two sub-problems that are alternately optimized at each iteration using successive convex approximation.

2) Numerical Results: This part discusses the impact of important parameters on the latency and the security performance of the FL network. For our simulations, we deploy the number of users $N = 50$, that is also used as in [3]. Since we aim to optimize the FL latency, the global accuracy is fixed at 10^{-3}. It is worth noting that the security performance obtained by the covert communication, i.e., the covert probability, is different from that obtained by the PLS, e.g., secrecy rate. Thus, it is not reasonable to compare the two approaches. Figure 4(a) shows the impact of the number of MDs N and the security threshold ϵ on FL latency. As observed, given ϵ, the FL latency increases as the number of devices N increases. The reason is that the same fixed bandwidth is allocated to more devices. This decreases the transmission rate of each device, thereby increasing FL latency. It can also be observed from Figure 4(a) that, as ϵ increases, FL latency decreases. The reason is that as ϵ increases, the security requirement $(1-\epsilon)$ decreases that allows the devices to transmit the models with higher transmit power. This leads to increases in SINR at the BS, thereby decreasing FL latency. Recall that when $(1-\epsilon)$ decreases, Willie can detect the transmissions of the devices more easily. As such, there is a trade-off between security performance and FL training efficiency.

Now, we discuss the impacts of the security threshold ϵ and the server's budget χ, FL latency and security performance. As shown in Figure 4(b), as ϵ increases, the covert probability of the FL network decreases. This is obvious since as ϵ increases (i.e., $(1-\epsilon)$ decreases), a lower covert probability is enough to satisfy the CC constraint. Figure 4(b) further shows that the covert probability remains almost constant over different budget settings by the FL server. The reason is that the covert probability depends on the ratio of the jamming power to the transmit power of the devices. As the budget varies, the jamming power bought from the jammer and the transmit power committed by the devices change together to satisfy the security requirement. Therefore, the covert probability remains unchanged over diverse budget values.

Nevertheless, varying the budget of the server leads to changes in FL latency. As shown in Figure 4(b), as we decrease the server's budget from $\chi = \$30$ to $\chi = \$10$, FL latency increases. The reason is that the server with a low budget can only buy a low amount of power from the jammer. The lower jamming power requires the devices to reduce their transmit power to satisfy the security requirement $(1-\epsilon)$ (i.e., to prevent Willie from detecting the transmissions from the MDs). This leads to decreases in SINR at the BS, thereby increasing FL latency. Under a higher budget setting (i.e., \$30), FL latency does not change significantly. The reason is that the server already finds an optimal jamming power that minimizes the FL latency while guaranteeing the security requirement, and it does not need to buy more power from the jammer. For this, the devices are not allowed to increase the transmit power due to the fixed security requirement. Thus, FL latency remains stable even when the budget is high.

IV. Conclusions and Future Directions

In this article, we present a vision towards the holistic and cost-effective protection of FLNs from attacks through a CC-based approach. We start by discussing existing security issues for distributed collaborative training of FL networks. We then review key existing CC techniques, and present a case study in which jammer-based CC is used to prevent an attacker from detecting the local model transmissions by the devices involved in an FL setting. The use of the AN signals leads to the increase of the FL latency. Thus, we have investigated the FL latency minimization problem subject to the CC constraint.

The numerical results is provided to evaluate the performance of the proposed approach. The covert communication approach can also be applied in centralized learning systems where the MDs share their training data rather than their locally trained model with the server, and the data transmissions are more privacy-critical.

For this emerging field of research, many interesting and challenging problems remain open:

- *Impact of multiple attackers:* There may be multiple attackers, and the server needs to prevent all of them from detecting the transmissions of the devices. This is challenging since the attackers may have different detection capabilities. One solution is to design the server to focus on defending against the worst-case attacker, i.e., the attacker that achieves the lowest covert rate.
- *Dynamic pricing of jamming power:* In this work, the jamming power price is fixed. In fact, the jammer can be self-interested and can dynamically set the price in different FL iterations to maximize its benefit. The server needs to account for time-varying prices to adapt the jamming power to avoid exceeding its budget.
- *Use of intelligent reflection surface (IRS):* To reduce the high cost for the jamming power, IRS can be deployed to enhance CC in the FL network. An IRS consists of reconfigurable reflecting elements that can reshape the phases, amplitudes, and reflecting the angles of the environmental signals. For this, the phase shifts of the IRS can be configured to maximize the SINR at the BS, subject to the CC requirement.

ACKNOWLEDGMENTS

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.02-2019.305, by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-019); the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme; Alibaba Group through Alibaba Innovative Research (AIR) Program and Alibaba-NTU Singapore Joint Research Institute (JRI); the Nanyang Assistant Professorship (NAP); the programme DesCartes; the RIE 2020 Advanced Manufacturing and Engineering (AME) Programmatic Fund (No. A20G8b0102), Singapore; Singapore Ministry of Education (MOE) Tier 1 (RG16/20); National Natural Science Foundation of China (NSFC) under grant No. 62102099; Postgraduate Innovation Foundation Project of Hebei Province of China under Grant CXZZBS2021137; and the China Scholarship Council (CSC). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

REFERENCES

[1] J. Tan, Y.-C. Liang, N. C. Luong, and D. Niyato, "Toward smart security enhancement of federated learning networks," *IEEE Network*, vol. 35, no. 1, pp. 340–347, Jan./Feb. 2020.

[2] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, "Toward an intelligent edge: Wireless communication meets machine learning," *IEEE communications magazine*, vol. 58, no. 1, pp. 19–25, Jan. 2020.

[3] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1935–1949, Mar. 2020.

[4] L. Lyu, H. Yu, X. Ma, L. Sun, J. Zhao, Q. Yang, and P. S. Yu, "Privacy and robustness in federated learning: Attacks and defenses," *preprint arXiv:2012.06337*, 2021.

[5] X. Lu, E. Hossain, T. Shafique, S. Feng, H. Jiang, and D. Niyato, "Intelligent reflecting surface enabled covert communications in wireless networks," *IEEE Network*, vol. 34, no. 5, pp. 148–155, Sep./Oct. 2020.

[6] X. Jiang, X. Chen, J. Tang, N. Zhao, X. Y. Zhang, D. Niyato, and K.-K. Wong, "Covert communication in UAV-assisted air-ground networks," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 190–197, Aug. 2021.

[7] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6193–6206, Sep. 2017.

[8] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE journal on selected areas in communications*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[9] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[10] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2031–2063, 3rd Quart. 2020.

[11] O. El Mouaatamid, M. Lahmer, and M. Belkasmi, "Internet of things security: Layered classification of attacks and possible countermeasures," *electronic journal of information technology*, no. 9, 2016.

[12] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[13] Y. Aono *et al.*, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1333–1345, May 2017.

[14] S. Vadlamani, B. Eksioglu, H. Medal, and A. Nandi, "Jamming attacks on wireless networks: A taxonomic survey," *International Journal of Production Economics*, vol. 172, pp. 76–94, 2016.

[15] T.-X. Zheng, H.-M. Wang, D. W. K. Ng, and J. Yuan, "Multi-antenna covert communications in random wireless networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1974–1987, Mar. 2019.

PLACE PHOTO HERE

Yuan-ai Xie received the B.S. degree in automation from North China University of Science and Technology, Tangshan, China, in 2016. He is currently pursuing the Ph.D. degree in control science and engineering in Yanshan University, Qinhuangdao, China. His current research interests include wireless resource optimization in vehicular networks and covert communications.

PLACE PHOTO HERE

Jiawen Kang received the M.S. degree and the Ph.D. degree from the Guangdong University of Technology, China, in 2015 and 2018. He is currently a full professor at the Guangdong University of Technology. He was a postdoc at Nanyang Technological University from 2018 to 2021, Singapore. His research interests mainly focus on blockchain, security, and privacy protection in wireless communications and networking.

PLACE PHOTO HERE

Dusit Niyato [M'09, SM'15, F'17] is currently a professor in the School of Computer Science and Engineering, at Nanyang Technological University, Singapore. He received B.Eng. from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand in 1999 and Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada in 2008. His research interests are in the areas of Internet of Things (IoT), machine learning, and incentive mechanism design.

PLACE PHOTO HERE

Nguyen Thi Thanh Van received the B.S. degree from Hanoi University of Industry, Hanoi, Vietnam, in 2010, the M.S. degree from Le Quy Don Technical University, Hanoi, Vietnam, in 2014. Her current research interest includes game theory, machine learning, and optimization techniques in communication networks.

PLACE PHOTO HERE

Nguyen Cong Luong received the M.S. degree in electronic and telecommunication engineering from the Hanoi University of Science and Technology (HUST). His research interest includes the next generation networks.

PLACE PHOTO HERE

Zhixin Liu received his B.S., M.S., and Ph.D. degrees in control theory and engineering from Yanshan University, Qinhuangdao, China, in 2000, 2003, and 2006, respectively. He is currently a professor with the Department of Automation, School of Electrical Engineering, Yanshan University, China. His current research interests include resource allocation in cognitive radio networks and vehicular networks.

PLACE PHOTO HERE

Han Yu is currently a Nanyang assistant professor with the School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU). His research focuses on the ethics of artificial intelligence and federated learning. He received the B.Eng. (Hons.) and Ph.D. degrees from the SCSE, NTU, Singapore, in 2007 and 2014, respectively.