International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

SCA Sybil-Based Collusion Attacks of IOT Data Poisoning in Federated Learning

P.Swapna[1] and Deena Babu Mandru[2]
[1]Assistant Professor, Department of Information Technology, MREC (A), Hyderabad-500100
[2]Professor, Department of Information Technology, MREC (A), Hyderabad-500100

Abstract
With the massive amounts of data generated by Industrial Internet of Things (IIoT) devices at all moments, federated learning (FL) enables these distributed distrusted devices to collaborate to build machine learning model while maintaining data privacy. However, malicious participants still launch malicious attacks against the security vulnerabilities during model aggregation. This paper is the first to propose sybil-based collusion attacks (SCA) in the IIoT-FL system for the vulnerabilities mentioned above. The malicious participants use label flipping attacks to complete local poisoning training. Meanwhile, they can virtualize multiple sybil nodes to make the local poisoning models aggregated with the greatest possibility during aggregation. They focus on making the joint model misclassify the selected attack class samples during the testing phase, while other non-attack classes kept the main task accuracy similar to the nonpoisoned state. Exhaustive experimental analysis demonstrates that our SCA has superior performance on multiple aspects than the state-of-the-art.

Keywords: Collision, IOT, Federated Learning and sybil-based collusion attacks

I. INTRODUCTION

WITH the fast development of industry 4.0 and the widespread popularity of industrial Internet of Things (IIOT) applications makes applications such as smart transportation and smart healthcare thrive and also makes the data generated by the industrial devices exponentially grow. Such as autonomous driving technology [1], it needs to train all data generated by sensor and camera devices to build a stable joint model to identify road conditions. And the distributed IIOT devices can generate a large amount of data in a short time [2]. In order to take into account the efficiency of processing big data and protect the privacy of clients. A novel machine learning paradigm named federated learning (FL) [3] was proposed, which is a new solution based on distributed training to alleviate the performance bottleneck and privacy risk caused by centralized processing. Traditional machine learning methods [4] usually store and run these data centrally, which will generate considerable computational and communication overhead in involving millions of mobile devices or massive data. This makes it unacceptable for sensitive IIOT applications (e.g., autonomous driving, intelligent robots, smart medical) that require real-time data transmission [5]. In addition, relying on centralized storage will cause a huge risk of private leakage [6]. Generally, when FL performs the collaborative training process of multiple distributed participants (e.g., IIOT devices), the sensitive information and private data of each client are kept locally [7]. FL has demonstrated excellent performance in the distributed execution process, while ensuring the

privacy of participants by performing independent local training and model updates, so as to implement collaborative calculating in a joint environment that includes malicious participants. This also makes FL attract much attention in many fields including smart healthcare [8] [9], smart feature prediction [10], and Internet of Things in smart homes [11] [12].

The IIOT represents a distributed network composed of intelligent and highly interconnected industrial devices, eac device can act as an FL participant to participate in training and updating [13]. FL improves the performance of the model for IIOT applications through continuous iterative training, and finally obtains a stable global model when the iteration reaches convergence. However, FL greatly exposes its weaknesses to malicious adversaries during the process of performing training [14]. Malicious adversaries can obtain the information of the global model in each round and upload malicious parameters or perform a small part of the beneficial contribution for collaborative training while avoiding anomaly detection as much as possible. For instance, malicious adversaries use contaminated data for training locally [15] [16], or tamper and prune local models for poisoning aggregation [17] [18].

The existing works [12] [19] have shown that controlling more malicious IIOT devices or using more direct poisoning attacks during the execution of FL is more destructive to the global model. Due to network, communication, power, and other issues in a heterogeneous federated environment, many IIOT devices are at risk of offline. Malicious participant will virtualize multiple malicious nodes in this unstable communication network. With more significant damage to the construction of the shared global model, this byzantine fault tolerance [20] problem usually uses the technology of fusion sybil-based attacks. In addition, in the process of malicious participants performing poisoning attacks, they usually use mislabeled samples for training or upload the poisoned models to the central server for aggregation. Compared with the independent attacks by a single malicious participant, the collusion attacks by multiple malicious participants have a higher attack success rate and can better obscure their attack behavior. Meanwhile, due to the characteristics of data privacy protection, the central server cannot verify the local data of all participants, and the parameter transmission process of all participants is anonymous, which provides more possibilities for malicious participants to launch malicious attacks.

Therefore, in order to better focus on the implementation of poisoning attacks in the IIOT-FL system, in this work, we introduce an efficient sybil-based collusion attacks (SCA) scheme. We represent the malicious IIOT device as a malicious participant in our system. Precisely, first, in the FL computing environment that we set, all the participants can only control local data and cannot access the data of other participants. This enables them to better manipulate local data for poisoning training without being

detected. The most commonly used data poisoning methods are backdoor poisoning attacks [15] and label flipping attacks [16]. This paper uses label flipping attacks to conduct poisoning training on the massive data generated by IIOT devices, aiming to make the global model misclassify the selected attack class samples. However, the attack effect achieved by such an attack is insufficient. Second, we use the cloning properties of sybil that all sybil nodes virtualized by malicious participants will perform the same malicious operations during the training process and have equal attack influence. We consider making the malicious model has a higher probability to be aggregated during FL aggregation. Finally, we collude with all malicious participants to launch the collusion attacks, aiming to replace the global model using the poisoning model. Meanwhile, such collusion attacks can better obscure their attack behavior. We utilize Fashion MNIST and CIFAR-10 datasets to represent the data generated by IIOT devices and conduct experiments. In summary, our contributions to this work are mainly in four-fold as below.

• We explore sybil-based collusion attacks of IIOT data poisoning for the IIOT-FL application, and implement poisoning training and model collusion attacks in this
IIOT-FL system.

We make minimal malicious assumptions for malicious adversaries and integrate the label flipping poisoning attacks to make the global model misclassify the selected attack class samples while maintaining the main task accuracy of other non-attack classes.

We further propose an efficient sybil-based collusion attacks (SCA) method, which aims to make the poisoning collusion models to be aggregated with greater probability during aggregation, and successfully obscure their attack behavior.

We utilize F-MNIST and CIFAR-10 datasets to represent the data generated by IIOT devices. Exhaustive experimental analysis demonstrates that our SCA has superior performance than the state-of-the-art.

II. LITRATURE SURVEY

Xie et al. [22] manipulated a subset of training data by injecting adversarial triggers to perform the wrong prediction on images embedded with triggers in a distributed heterogeneous dataset. Sun et al. [23] injected backdoor tasks into a part of the images to damage the global model's performance on the target task. Although it has a high attack success rate, it can cause much overhead to inject backdoor triggers into large-scale training samples. In addition, the goal of our attack is to misclassify the selected attack class samples. So in this work, we use the label flipping poisoning attacks. Malicious adversaries can perform label flipping attacks without conducting parameter interaction, changing the FL architecture, and pre-training. They use the dirty data with the wrong label for training locally. This attack method is both concealed and direct.

Jiang et al. [25] proposed a sybilbased attacks method. Sybil clients compromised the infected device to

update the poisoning model directly. They proved their effectiveness on several advanced defense methods, while also slowing down the convergence of the global model. Fung et al. [26] also designed a novel sybil-based attacks technology, it has shown the effectiveness on multiple recent distributed machine learning fault tolerance protocols. The sybil attacks also showed an excellent attack effect in IoT applications [27]. Although they have shown reliability in the attack effect, the drift gradient of their local poisoning model is very easy to detect and remove. In this paper, we integrate the sybil-based collusion attacks technology to make the local poisoning model have a higher possibility of aggregation and help malicious participants better obscure the attack behavior.

Taheri et al. [28] proposed two dynamic poisoning attack strategies that integrate Generative Adversarial Network (GAN) and Federated Generative Adversarial Network (FedGAN) on the side of the participants, and evaluated them on IIoT applications. Lim et al. [29] studied the collusion attacks between dishonest participants and the server. The malicious participant uploads the poisoning model during the aggregation stage, and the server also leaks the parameters of other participants to the malicious participant. They aim to achieve the purpose of reducing the global model's performance while analyzing the local model of other participants to avoid anomaly detection [30] during the poisoning process.

The system is not implemented the use the cloning properties of sybil that all sybil nodes virtualized by malicious participants will perform the same malicious operations during the training process and have equal attack influence.

The system is not implemented SCA on IIoT-FL model.

III. PROPOSED METHODOLOGY

The proposed system explores sybil-based collusion attacks of IIoT data poisoning for the IIoT-FL application, and implement poisoning training and model collusion attacks in this IIoT-FL system.

The proposed system makes minimal malicious assumptions for malicious adversaries and integrate the label flipping poisoning attacks to make the global model misclassify the selected attack class samples while maintaining the main task accuracy of other non-attack classes.

The proposed system further propose an efficient sybil-based collusion attacks (SCA) method, which aims to make the poisoning collusion models to be aggregated with greater probability during aggregation, and successfully obscure their attack behavior.

The proposed system utilizes F-MNIST and CIFAR-10 datasets to represent the data generated by IIoT devices. Exhaustive experimental analysis demonstrates that our SCA has superior performance than the state-of-the-art.

The system is implemented SCA Based on Label Flipping Poisoning Attacks which is more secured and safe.

In the proposed system, the system is implemented scenario that the malicious adversary uses the label flipping strategy

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

to train the poisoning data locally and collude with other poisoning models.

IV. IMPLEMENTATION

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

Login, Browse Network Datasets and Train & Test Data Sets, View Trained and Tested Network Datasets Accuracy in Bar Chart, View Trained and Tested Network Datasets Accuracy Results, View Prediction Of Sybil based Collusion Attack Status, View Sybil based Collusion Attack Status Ratio, Download Predicted Data Sets , View Sybil based Collusion Attack Status Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

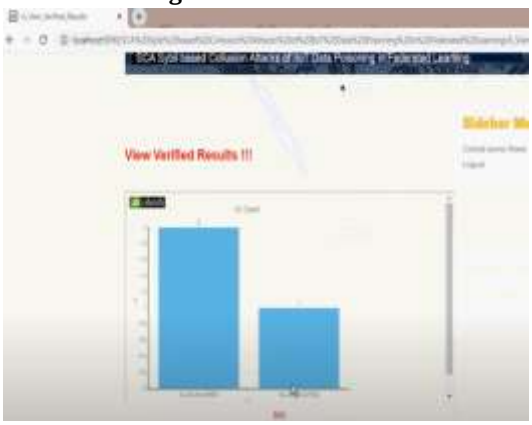
Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT SYBYL BASED COLLUSION ATTACK STATUS, VIEW YOUR PROFILE.

Fig.1. Home page.

Fig.2. Server details.

Fig.3. Vehicle details.



International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Fig.4. Output results.

Fig.5. User details.

Fig.6. Vehicle data set.

V. CONCLUSION

This paper analyzed the security vulnerabilities of joint training in the IIOT-FL system, then proposed a sybil-based collusion attacks (SCA) approach for the vulnerabilities. Meanwhile, we also gave further details on the execution of related algorithms, model architecture, and analysis of the effectiveness of the experiment. In this work, malicious participants in our federated system can virtualize multiple Sybil nodes and perform malicious collusion attacks. The purpose is to make the local poisoning model be aggregated with a greater possibility. They aim to make the samples of the selected attack class be misclassified, while other non-attack classes maintain similar accuracy as before. Compared with the state of- the-art, our SCA can achieve a more substantial attack effect under the condition of fewer malicious participants performing collusion, and can successfully obscure their attack behavior. Extensive experimental results show that our SCA has a more robust attack performance on several evaluation metrics.

REFERENCES

[1] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyatoand H. V. Poor, "Federated learning for industrial internet of things infuture industries," IEEE Wireless communications magazine, 2021.

[2] P. Zhang, C. Wang, C. Jiang, and Z. Han. "Deep reinforcement learningassisted federated learning algorithm for data management of IIoT,"IEEE Transactions on Industrial Informatics (TII), 2021.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas,"Communication-efficient learning of deep networks from decentralizeddata," in Proceedings of the 20th International Conference on ArtificialIntelligence and Statistics (AISTATS), 2017, pp. 1273-1282.

[4] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collectionfor machine learning: A big data - AI integration perspective," IEEETransactions on Knowledge and Data Engineering (TKDE), vol. 33, no.4, pp. 1328-1347, 2021.

[5] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang, and Y. Liang, "Blockchainenabledfederated learning data protection aggregation scheme withdifferential privacy and homomorphic encryption in IIoT," IEEE Transactionson Industrial Informatics (TII), 2021.

[6] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A.

Dehghantanha,and G. Srivastava, "A survey on security and privacy of federatedlearning," ELSEVIER Future Generation Computer Systems (FGCS),vol. 115, pp. 619-640, 2021.

[7] T. Li, A. K. Sahu, A. Talwalker, and V. Smith, "Federated learning:Challenges, methods, and future directions," IEEE Signal ProcessingMagazine, vol. 37, no. 3, pp. 50-60, 2020.

[8] W. S. Zhang, T. Zhou, Q. H. Lu, X. Wang, C. S. Zhu, H. Y. Sun, Z.P. Wang, S. K. Lo, and F. Y. Wang, "Dynamic fusion-based federatedlearning for COVID-19 detection," IEEE Internet of Things Journal(IoTJ), 2021.

[9] M. Parimala, M. S. Swarna, P. V. Quoc, D. Kapal, M. Praveen,T. Gadekallu, and T. H. Thien, "Fusion of federated learning andindustrial internet of things: A survey," arXiv preprint arXiv:2101.00798,2021.

[10] M. X. Duan, K. L. Li, A. J. Ouyang, K. N. Win, K. Q. Li and Q. Tian,"EGroupNet: A feature-enhanced network for age estimation with novelage group schemes," ACM Transactions on Multimedia Computing,Communications, and Applications (TOMCCAP), vol. 16, no. 2, 2020.