*Rethinking Data Integrity in Federated Learning : Are we ready ?*

1Shreyas R. Dixit   
 *UG Student, E&TC Engg., Vishwakarma Institute of Information Technology , Pune*

Email:shreyas.22111142@viit.ac.in

2Parikshit N. Mahalle,   
Senior Member IEEE

*AI &DS Engg., Vishwakarma Institute of Information Technology , Pune* , Email:parikshit.mahalle@viit.ac.in

2Gitanjali R Shinde,   
*Computer Engg., Vishwakarma Institute of Information Technology, Pune*

### *Email:*gitanjali*.*shinde*@viit.ac.in*

***Abstract*—Federated Learning allows to train high level machine learning models with user’s data without the data leaving the edge device. This allows Federated Learning (FL) models to create better personalized results for the user’s community.As Federated Learning is fairly new, it is susceptible to data leakage or even data manipulation by attackers.The architecture of FL is an Iterative process, hence even a small data manipulation can lead to exploding errors.The main aim of Federated Learning is to keep user’s data on the edge device. This is achieved by sending the base Machine Learning model to the edge device and training the models on the user’s data, then weights and biases of these trained models are sent to the “Access Layer” at the server end which collects these parameters from all the devices on the FL network. By this the data is not actually sent to the server but the model learns the weights & biases.Considering the high risk of data manipulation through cyberattacks, the approach of the study in this paper is to learn about data integrity in transit in Federated Learning wherein the different types of cyberattacks have been explored and activity attack modeling has been done on federated learning architecture. Conclusive remarks have been presented along with different characteristic and behavioral properties of various cyber attacks**

***Keywords—Federated Learning , cyber attacks, Data* Integrity *in transit.***

# Introduction

Normally a Machine Learning model is for “Centralized Data” which means that all the data as well as the model are on the same device/server (location). But when there are multiple data locations we send the data to the location where the model is trained and the output is sent to the user. In Federated Learning we try to flip the logic and instead of sending the data to the cloud, we try to send the model to the device & train these models locally. This ensures the data never leaves the device. After this we send the updated model to the cloud which only includes the updated weights and biases. After this, the models are combined using the avg. method or aggregate method.

The main aim of Federated learning is to provide maximum privacy, this is accomplished by keeping the data on the device. The models are trained on the device & only the updated weights are sent back to the cloud. This allows us to maintain data privacy while having data from multiple different locations. Federated Learning also provides us with great quality results no matter the data size you provide. This is because it aggregates all the weights & biases which the individual devices send.

The most important part about Federated Learning is that it keeps user’s data private.This allows user’s edge device to learn & adapt to the user’s data without sending the user’s data out of the edge device.This means user’s smartphone does not send any personal information to the central server. It only sends the updated parameters of the machine learning model.This ensures that the user’s data remains on the device all the time. This provides user’s with better results while keeping their data safe with them , rather than worrying about their data being on the cloud at the time where cyber-attacks have been increasing steadily.

As we can see Federated Learning is providing us with great levels of data privacy and state of art machine learning & deep learning models. But while doing this there are some places where the architecture of federated learning is falling short. The main motivation behind this paper is to understand the vulnerabilities in the architecture of federated learning using activity based attack modeling rethink about deploying large scale machine learning models in federated learning architecture as data integrity in this architecture is not perfectly secure.

# Literature Review

Yang el.al.[1] has explained the concept of Federated Learning as an Architecture which allows to collaboratively train & use a shared prediction model while keeping all the local training data private. T. Li el.al.[2] goes over the challenges and limitations faced by the Federated Learning architecture. Bonawitz. K el. al. [3] paper explains the concept of Distributed Federated Learning over the traditional Centralised Federated Learning. Kairouz, P. el.al [4] gives us a broad overview of the existing open problem in the domain of Federated Learning.FL embodies the principles of focused data collection and minimization, and can mitigate many of the systemic privacy risks and costs resulting from traditional, centralized machine learning and data science approaches. Zhang,C el.al [5] describes the characteristics and the current practical application of federated learning & its working. Lyu, L [6] dives into the possible threats in the Federated Learning Architecture & FL protocol design has been shown to exhibit vulnerabilities which can be exploited by adversaries both within and without the system to compromise data privacy. Poisoning attacks and inference attacks, this paper provides an accessible review of this important topic. Rodríguez-Barroso el.al [7] Describes new challenges like adversarial attacks in Federated Learning and define proper guide lines for defence against these attacks. Mahalle P el.al. [11] Talks about the protection of data and privacy in IOT devices and proposes a secure model for IOT devices. Mahalle P[13]In this paper studies about various upcoming security threats and vulnerabilities in the IoT as increasing number of devices are getting connected to the intternet. The paper also discussed about new types of cyber-threats & concludes with mitigation stratigies for trust attacks. [14] Dives into different types of Cyber-attacks on IOT based deives like Sybil , Man in the Middle , DOS , Ransomware ,etc. . It also provides mitigation approaches for these cyber attacks ,with activity attack modeling of each of these cyberattacks.

# Gap Analysis

Federated Learning architectures are used in a lot of applications such as next word suggestion on keyboard , product recommendation , etc. Even though a lot of work has been carried out in the domain of Federated Learning, Activity based attack modeling has not yet been done on Federated Learning Architecture. Activity Attack Modeling helps find vulnerabilities of the Architecture which helps us to decide from different mitigation approches for these threats, while helping us understand how these attacks work.

# Threat Analysis

Federated Learning is based on the concept of sending copies of the basic model to suitable devices & Training the model on the device & sending back the updated weights to the cloud for creating the new Updated model. This process is carried on over 1000’s iterations.

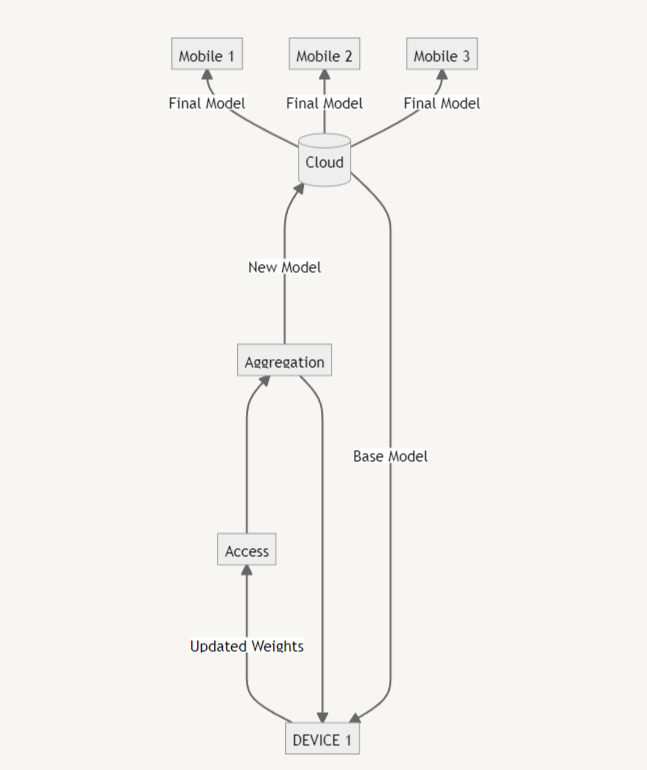


Figure1 : Federated Learning Architecture.

## First Basic Model from the Cloud is sent to the Device.

## Then the device trains the model on its personal data.

## The Updated weights of the model are then sent to the Access point

## These Access points then send all the Updated weights on to the Aggregation Layer

## In the Aggregation Layer, the weights from Multiple devices are combined using the Weighted Sum method, this produces new Weights and Biases which are then used to create an Efficient Model

## This model is then sent back to the devices as Base Model and the process is repeated over 1000s of Iteration until we get a model with accuracy high enough.

## This is then sent to the Cloud Storage from where it is Deployed to all the other devices.

In this full process, we assume that the data is safe while it is in the Device, Access point, Aggregation Layer, and Cloud.

The main area of concern is the transit where the data is sent from the Device to Access Point, or Access point to Aggregation Layer, or any transit in the Architecture. As only the weights are being sent as the data from layer to layer, there is a good probability that the data can be tampered with and its Integrity is lost. Even if a very small amount of data has been changed it will affect the overall model as this is an iterative process. The error in the calculations will go on increasing. This is similar to the Problem of exploding gradients faced in LSTMs.

There are Multiple Types of Threats that can occur in this transit, some of which are listed below.

* Interruption: Stops data transfer
* Interception: Logging Data
* Fabrication: Creating Data
* Modification: Changing Data

# Activity based Attack Modeling

This level represents the Device to Access Layer transit of data. We have tried to model the type of attacks that can occur in this layer and what loss in data integrity we can face. Below are 2 types of attacks that we can conduct on the transit of data from the Device to the Access Layer

Man in the Middle

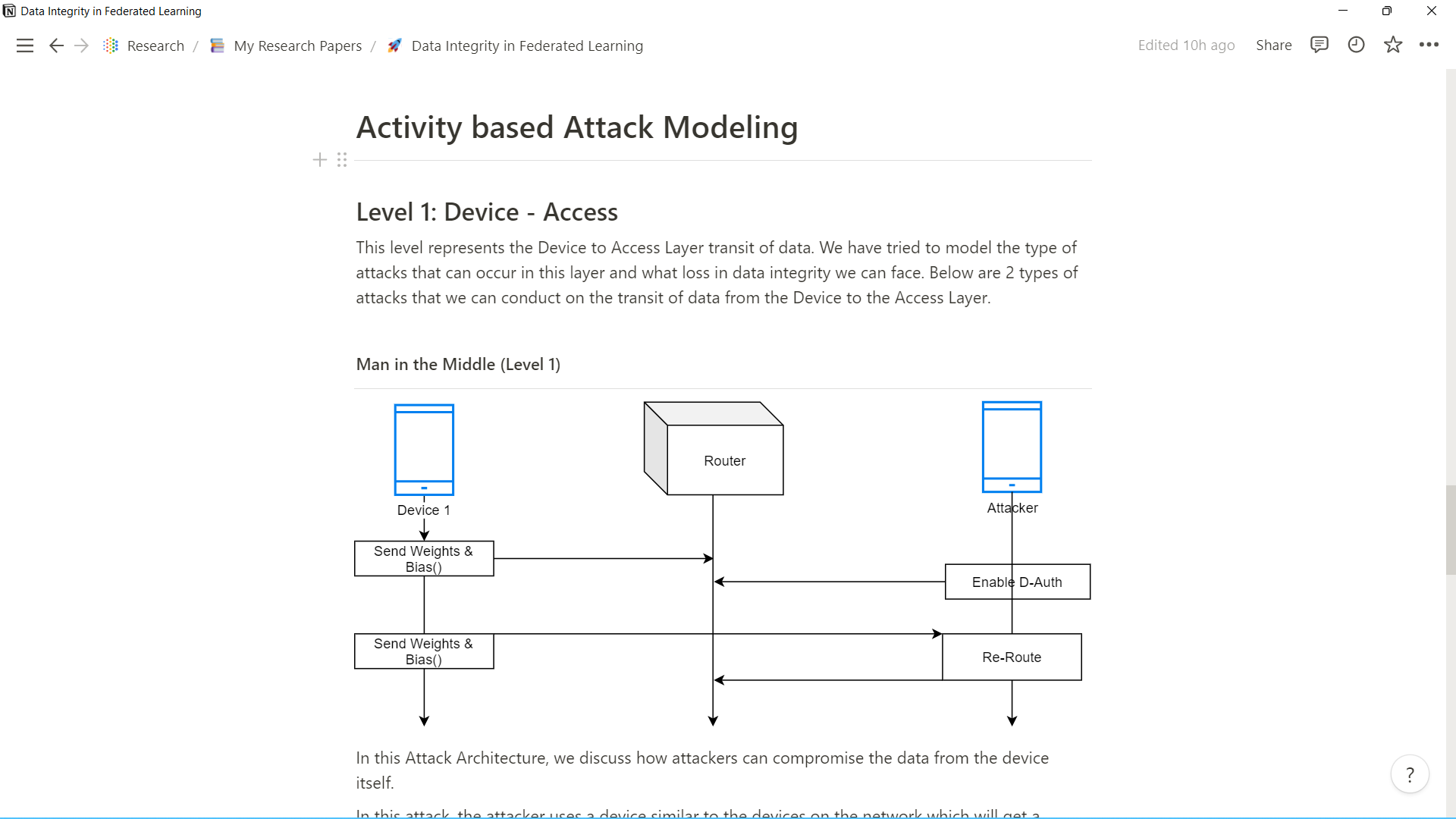


Figure 2 : Man in the Middle Activity Attack Modeling.

Figure 2 depicts the attack modeling of Man in the Middle attack. Below discussed attack architecture indicates attackers can compromise data from the device itself.

In this attack, the attacker uses a device similar to the devices on the network which will get a request from the cloud to train the Federated Model on its own data. The attacker device will work as a normal device in this time frame. But as the devices start to send in the model Weights and Bias the attackers send in an Enable D-Auth command which allows the attacker to redirect the entire data to flow from the attacker's device. This allows the attacker to not only view the data from other devices on the training network but also allows to change the data in between.

Sniffing

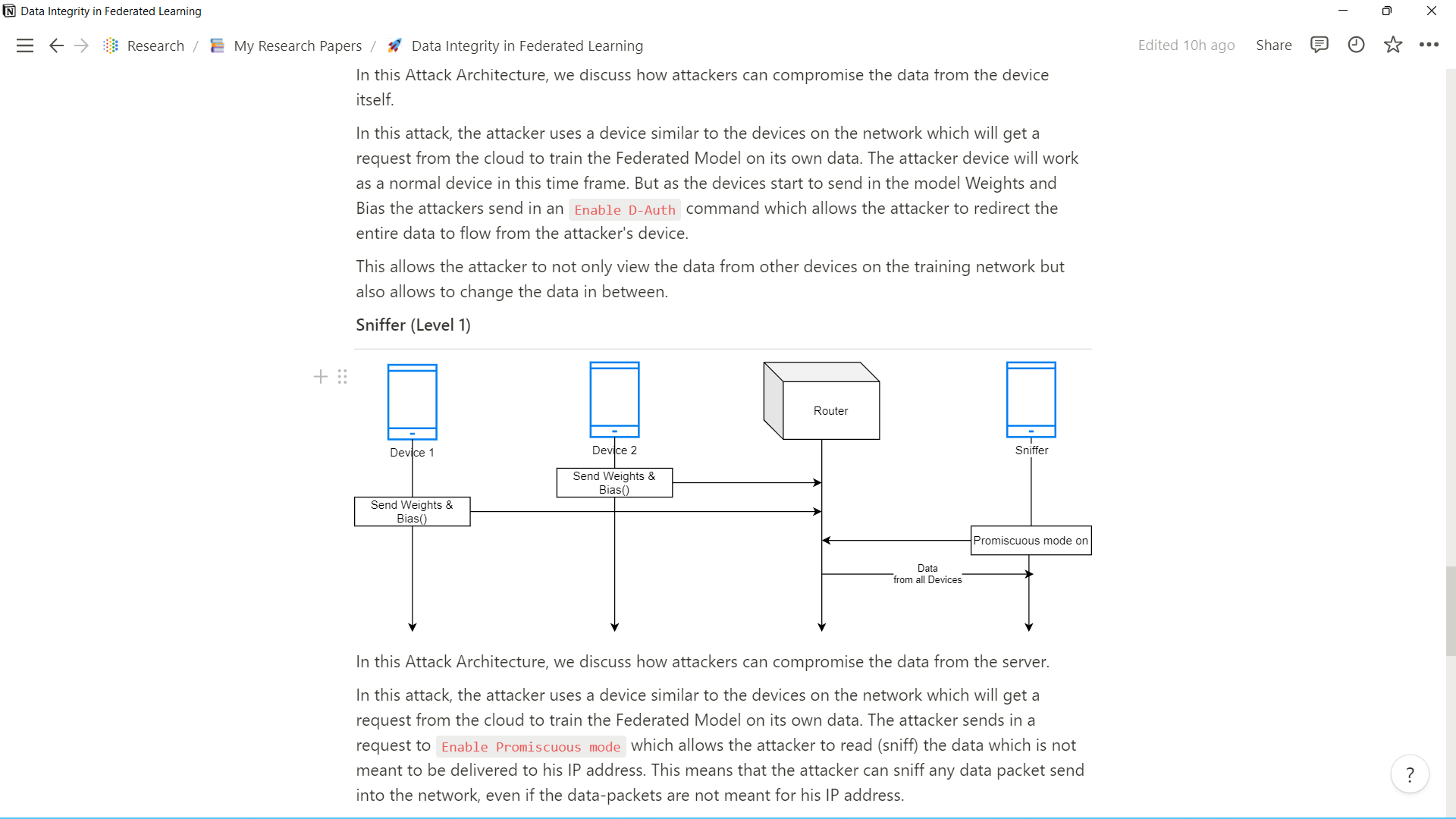


Figure 3 : Sniffing Attack Activity Modeling

Figure 3. shows the sniffing attack modeling.

In this Attack Architecture below, it is described how attackers can compromise the data from the server.

In this attack, the attacker uses a device similar to the devices on the network which will get a request from the cloud to train the Federated Model on its own data. The attacker sends in a request to Enable Promiscuous mode which allows the attacker to read (sniff) the data which is not meant to be delivered to his IP address. This means that the attacker can sniff any data packet sent into the network, even if the data-packets are not meant for his IP address.

Even though we are sending only Weights and Biases on the model, this creates a possibility of re-engineering the data on which the model was trained.

Sybil

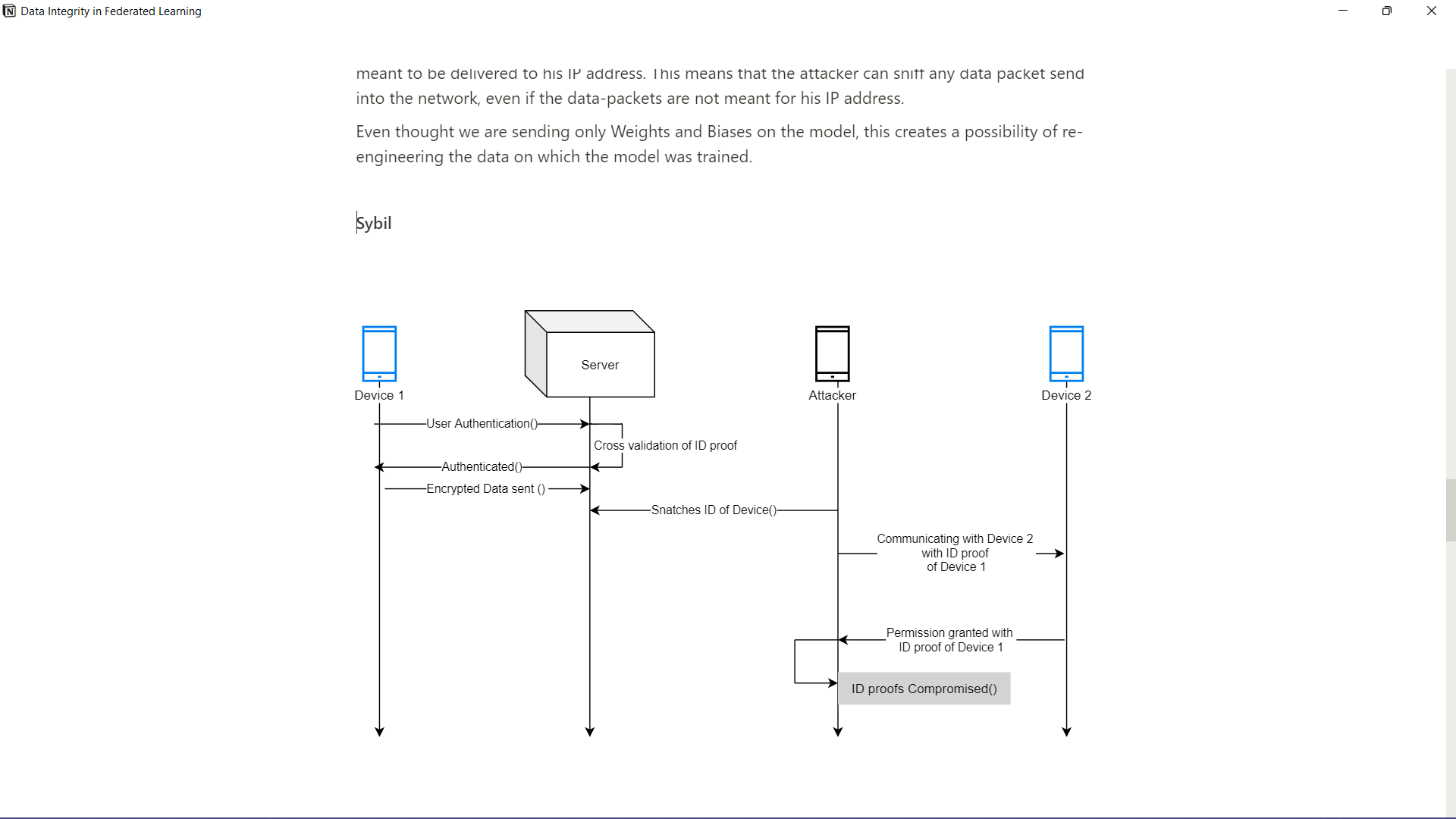


Figure 4 : Sybil Activity Attack Modeling

In this Attack Architecture below , studies show how attackers can take identity of another device on the network and extract information from rest of the devices.

Each device has some identity proof , in this attack the attacker communicates with the server to snatch ID of a device in the network.The attacker then creates a duplicate identity of that device and communicates with other devices on behalf of the compromised device identity.This allows the attacker to retrieve data from other devices. In this type of attack the user is present at multiple places at the same time. This can be done using Identity spoofing.

# Issues & Challenges

The conducted study suggests that data integrity in Federated Learning architecture is a part which can be heavily exploited, hence revisiting data integrity is very important before deployment.

Some of the challenges posed by this architecture are :

1. Distributed Architecture : Federated Learning uses Distributed Machine Architecture wherein there are multiple devices which need to be maintained & inspected regularly for malicious attackers trying to join the Federated Learning network. This allows the attacker to carry out an attack like “man in the middle.”

2. Decentralized Operation : The majority of Federated Learning which is currently being used is focused on Centralized Operation. This means that all the model weights & biases go to a server, where all these weights are then computed to generate a weighted sum. This weighted sum is then sent to the Aggregation layer which creates the model and sends the model to the cloud platform. In the decentralized approach weights & biases from all the devices will be sent to each other and then weighted sum from each device will be sent to the server to take a Fed-Average.

3. Heterogeneity : The data which FL models train on contains a wide variety of range as each user is unique and hence each user’s data is unique. This becomes a great challenge when combining the model weights & biases and getting a generalized model for deployment.

4. Resource Constraints : Even though we select the devices on which the FL models are to be trained , not all devices have the same memory , computational power , or battery. This causes delayed responses and low accuracy and interrupted models (half trained models).

# Possible Mitigation approaches

Security by Design : This is an approach to software and hardware development that seeks to make systems as free of vulnerabilities and impervious to attack as possible through such measures as continuous testing, authentication safeguards and adherence to best programming practices. It focuses on building security into the architecture at every layer.

Architectural Security : This is an approach which is mostly used in applications. It focuses on how to build security after the architecture is fully built.

Crypto Security : This is a fairly new but a very strong approach towards security. It revolves around the concept of Crypto-Encryption and Crypto-keys. The data which is to be sent is encoded using cryptography and only the receiver has the key for decoding this encryption.

Secure Aggregation : Secure aggregation is a popular protocol in privacy-preserving federated learning, which allows model aggregation without revealing the individual models in the clear

These are the possible mitigation approaches which could be addressed to mitigate the issues faced by Federated Learning Architecture.

# Concussion

All the above studies of activity based attack modeling regarding Federated Learning Architecture indicate that data integrity in transit in Federated Learning is not perfect and we may need to revisit & apply some mitigation techniques before we widely deploy the large scale machine learning models using federated learning architecture

##### References

1. Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T. and Yu, H., 2019. Federated learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 13(3), pp.1-207.
2. T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," in IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, May 2020, doi: 10.1109/MSP.2020.2975749.
3. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, B. and Van Overveldt, T., 2019. Towards federated learning at scale: System design. Proceedings of Machine Learning and Systems, 1, pp.374-388..
4. Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R. and D’Oliveira, R.G., 2021. Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), pp.1-210.
5. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W. and Gao, Y., 2021. A survey on federated learning. Knowledge-Based Systems, 216, p.106775.
6. Lyu, L., Yu, H. and Yang, Q., 2020. Threats to federated learning: A survey. arXiv preprint arXiv:2003.02133.
7. Rodríguez-Barroso, N., Jiménez-López, D., Luzón, M.V., Herrera, F. and Martínez-Cámara, E., 2023. Survey on federated learning threats: concepts, taxonomy on attacks and defences, experimental study and challenges. Information Fusion, 90, pp.148-173
8. Huba, D., Nguyen, J., Malik, K., Zhu, R., Rabbat, M., Yousefpour, A., Wu, C.J., Zhan, H., Ustinov, P., Srinivas, H. and Wang, K., 2022. Papaya: Practical, private, and scalable federated learning. Proceedings of Machine Learning and Systems, 4, pp.814-832.
9. Pappas, Christodoulos & Chatzopoulos, Dimitris & Lalis, Spyros & Vavalis, Manolis. (2021). IPLS: A Framework for Decentralized Federated Learning. 1-6. 10.23919/IFIPNetworking52078.2021.9472790.
10. Ergun, I., Sami, H.U. and Guler, B., 2021. Sparsified secure aggregation for privacy-preserving federated learning. arXiv preprint arXiv:2112.12872.
11. Babar, S., Mahalle, P., Stango, A., Prasad, N., Prasad, R. (2010). Proposed Security Model and Threat Taxonomy for the Internet of Things (IoT). In: Meghanathan, N., Boumerdassi, S., Chaki, N., Nagamalai, D. (eds) Recent Trends in Network Security and Applications. CNSA 2010. Communications in Computer and Information Science, vol 89. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-14478-3_42>
12. Handbook on ICT in Developing Countries: Next Generation ICT Technologies. N.p.: River Publishers, 2022.
13. Mahalle, P.N., Shinde, G.R. (2020). Trust Attacks in Internet of Things: A New Data-Centric Cybercrime on Enterprise Use Case. In: Vasudevan, H., Michalas, A., Shekokar, N., Narvekar, M. (eds) Advanced Computing Technologies and Applications. Algorithms for Intelligent Systems. Springer, Singapore. <https://doi.org/10.1007/978-981-15-3242-9_>
14. Railkar, P.N., Mahalle, P.N., Shinde, G.R. and Bhapkar, H.R., 2019. 3 Threat analysis and attack modeling for machine-to-machine communication toward Internet of things. The Internet of Everything, p.45.