



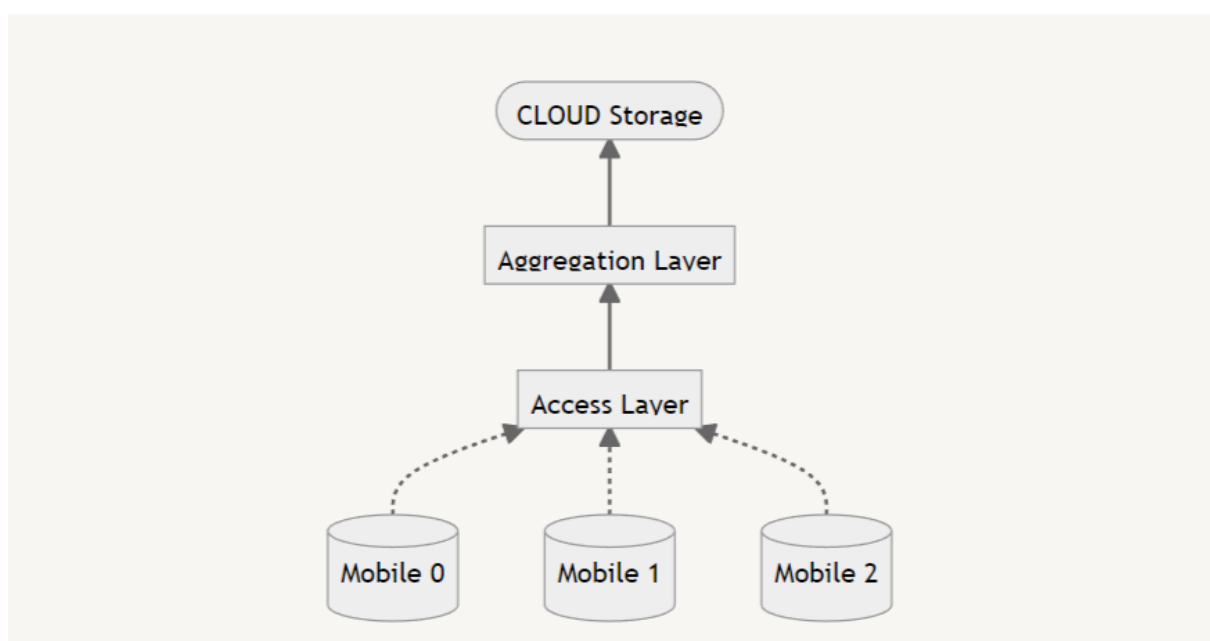
Data Integrity in Federated Learning

Introduction

What is Federated Learning?

Normally a Machine Learning model is for “Centralised 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 given 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.



Why Federated Learning ?

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 crucial reason for relying on Federated Learning is that it keeps things private on your part. This is because of how your machine or your mobile device continues to learn and adapt without the need to make use of the data sent over to it from the cloud. In that sense, your smartphone doesn't send over your personal data and information to a central server as only the updates on the phone are sent to it. That means that what is meant to stay on your mobile device will remain on your mobile device. This ability allows anyone to enjoy using their smartphone or other devices while minimizing the risk of their personal data and information being leaked online. Especially now that we live in an age where cyber-attacks are becoming more common as hackers and cyber-criminals are getting smarter and better at what they do.

Motivation

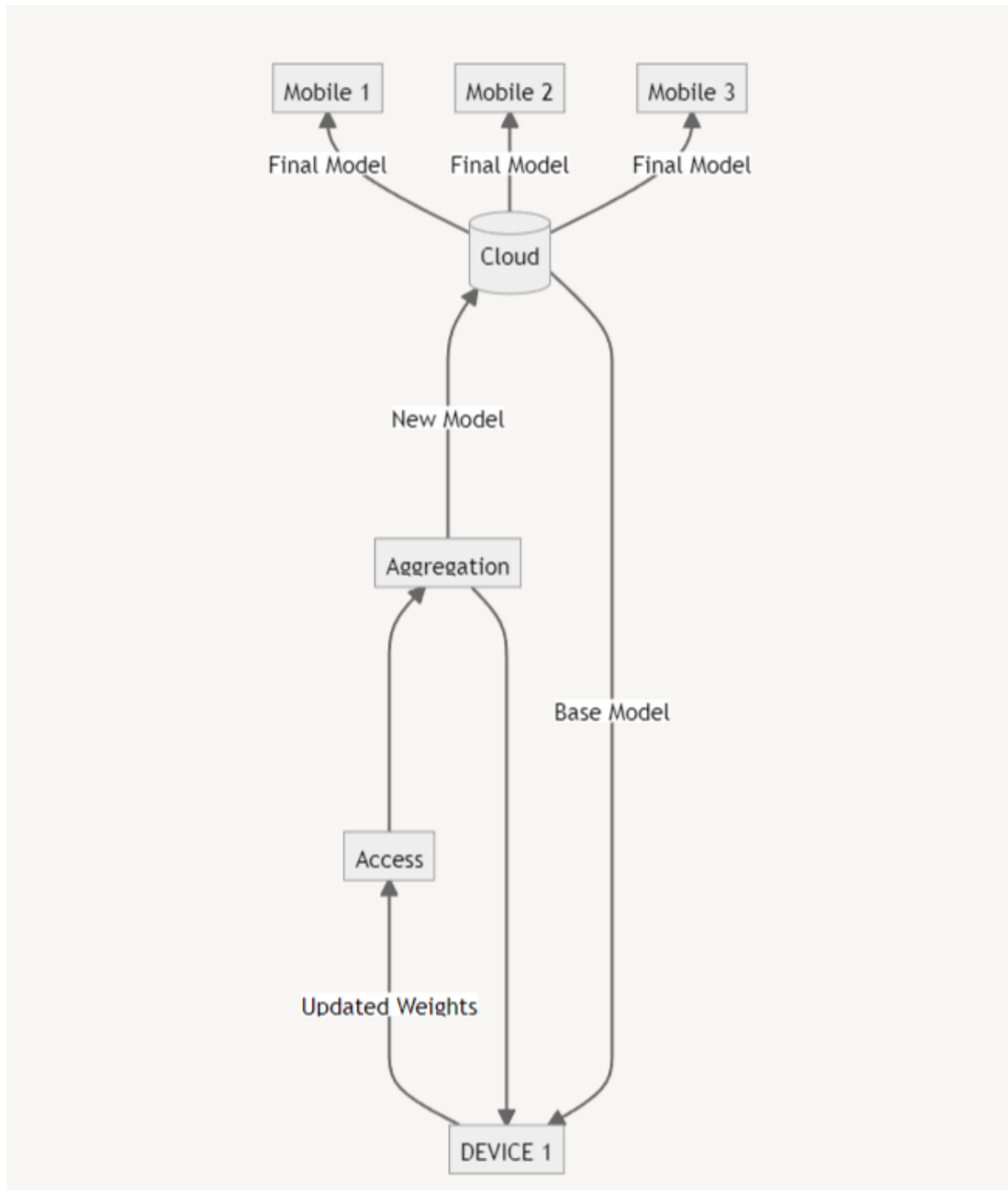
As we can see Federated Learning is providing us with great levels of Data Privacy and State of Art Machine Learning & Deep Learning Models. But while doing this there are some places where the Architecture of Federated Learning lacking. The main motivation behind this paper is to understand the Vulnerabilities in the Architecture of Federated and see how Tech Giants like Google, and Facebook, are handling these vulnerabilities & can they be improved further.

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 & send back the updated

weights to the cloud for creating the new Updated model. This process is carried on over 1000's iterations.

This Architecture is not as simple as it seems, rather not as secure. The data flow for the following is as follows :



- 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 less 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 Gradient faced in LSTMs.

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

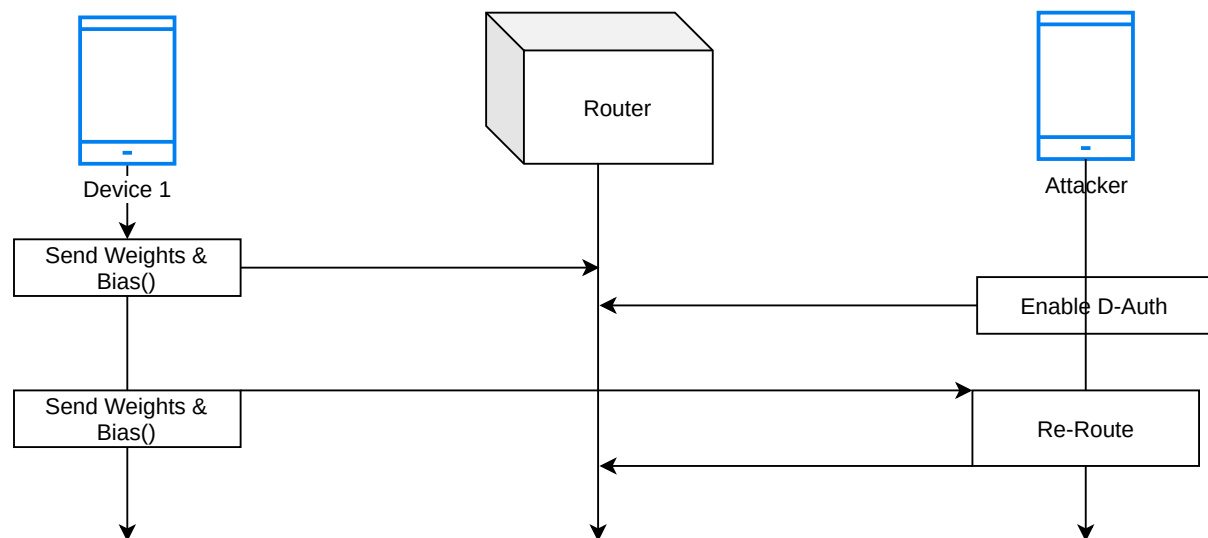
- Interruption: Stops data transfer
- Interception: Logging Data
- Fabrication: Creating Data
- Modification: Changing Data

Active Attack Modeling

Level 1: Device - Access

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 (Level 1)

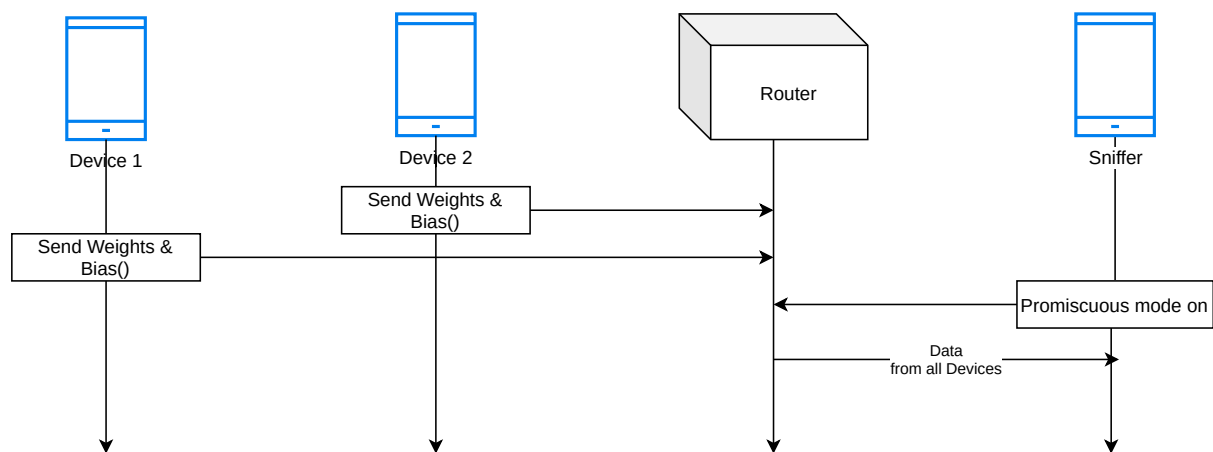


In this Attack Architecture, we discuss how attackers can compromise the data from the device itself.

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.

Sniffer (Level 1)



In this Attack Architecture, we discuss 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 send 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.

Level 2: Access - Aggregation

This level represents the Access to the Aggregation 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.

Man in the Middle (Level 2)

Sniffer (Level 2)

Level 3: Aggregation-Cloud

This level represents the Aggregation to Cloud 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.

Man in the Middle (Level 2)

Sniffer (Level 2)


Gap Analysis

Proposed Work

Results & Discussion

Conclusion

References

 [Resources](#)