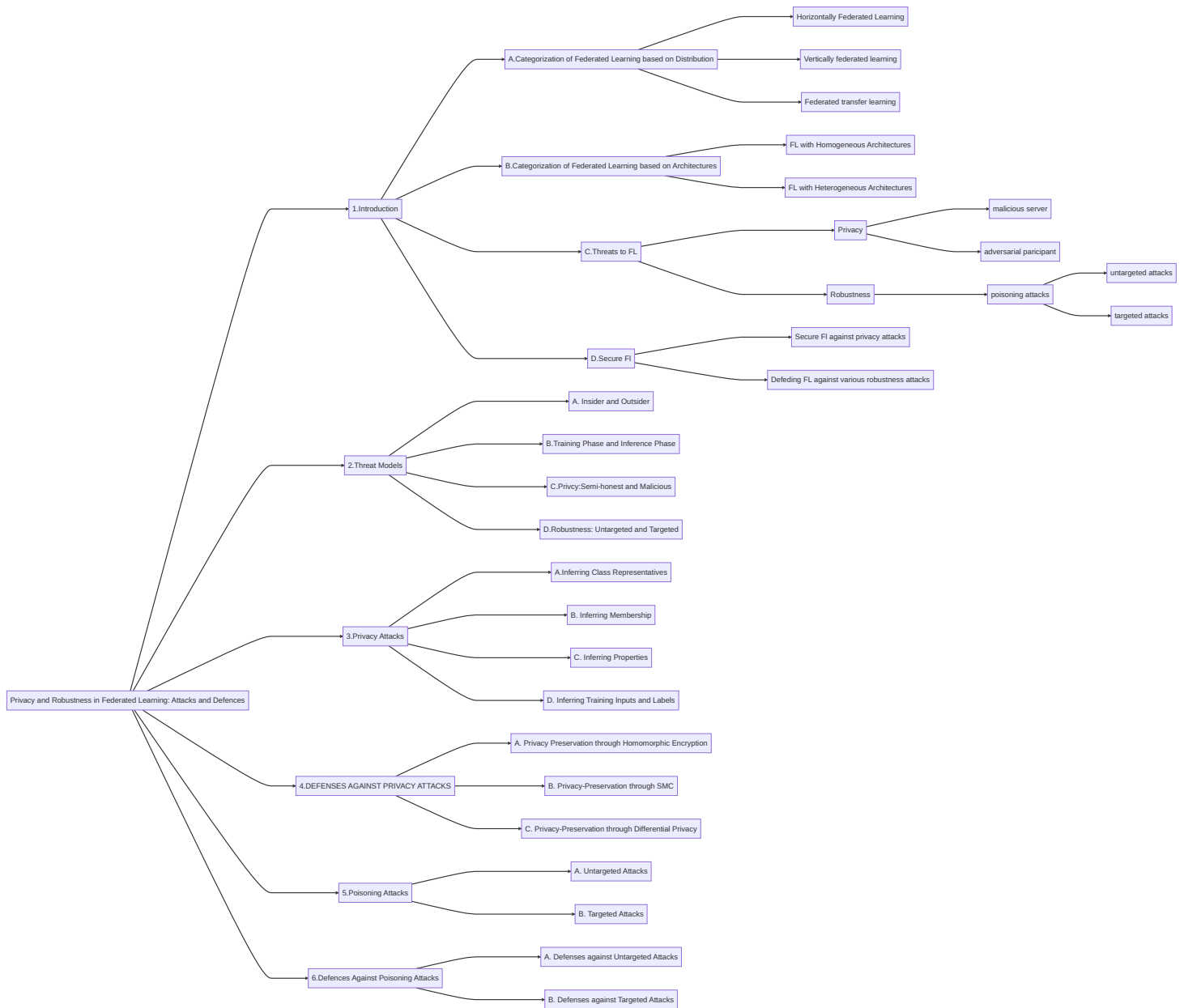


# A-Review-Of-Security-Issues-In-Federated-Learning

This is our first paper, which is called A Review Of Security Issues In Federated Learning.



Robust and Privacy-Preserving Collaborative Learning: A comprehensive Survey

2. System Overview:  
We systematically introduce different types of collaborative learning systems from various perspectives.

1. Introduction

A. Machine Learning Model

Data Parallelism

Model Parallelism

Pipelining

Hybrid Parallelism

B. Dimensions of Parallelism

C. Parameter Distribution

Centralized

Decentralized

D. Model Consistency

Synchronous

Asynchronous

State-Synchronous

E. Federated Learning

Compromise attacks aims to reduce or destroy the trained model performance by changing model parameters, which normally makes the shared model not converge to a satisfactory one during the training phase.

Backdoor attacks try to inject predefined malicious training samples, backdoors, into a victim model while maintaining the performance of the primary task. The backdoors would be activated if a input sample contains the injected triggers. Because of the secrecy of triggers, it is difficult to identify backdoor attacks as a backdoored model performs normally on normal samples.

3. Threat In Collaborative Learning:  
We summarize summarizes the privacy and integrity threats in collaborative learning

A. Integrity Threats

Compromise vs. Backdoor

Data Poisoning vs. Model Poisoning

Attackers can poison the training datasets of some participants with malicious samples with carefully crafted triggers

Attackers compromise some participants and completely control their behaviour during the training

B. Privacy Threats

Membership vs. Property vs. Sample

成员推理攻击确定记录是否在模型的训练数据集中

属性推理攻击旨在推断参与者训练数据的属性

攻击者在训练阶段获得模型更新时提取训练数据及其标签

Passive vs. Active

被动模式-不改变任何内容

主动模式-攻击者可以在训练过程中做任何事情

4. Integrity Attacks

A. Byzantine Attacks

B. Backdoor Attacks

Data Poisoning

Unclean label stand-alone backdoor the adversary introduces a number of miss-classified data samples into the training set, it poisons the training examples and changes their labels.

Model Poisoning

Clean label stand-alone backdoor the adversary cannot change the label of any training sample and preserves the labels of the poisoned samples

5. Integrity Defenses

Byzantine Defences

Statistic-based Inspection

Learning-based Inspection

Backdoor Defences

Data Inspection

Model Inspection

6. Privacy Attacks

A. Threat Model

B. Membership Inference

C. Property Inference

D. Sample Inference

7. Privacy Defenses

A. Differentially Private Collaborative Learning

B. Cryptographic Privacy-preserving Collaborative Learning

C. Practical Privacy-preserving Collaborative Learning

8. Hybrid Defenses And Beyond

A. Hybrid Defenses

B. Collaborative Adversarial Training

9. Open Problem

Non-IID or Noisy Scenarios in Byzantine Attacks and Defenses

Certified Backdoor Defenses

Privacy-performance Tradeoff in Differential Privacy

Basis Datasets in Property Inference Attacks

Performance Improvement in Sample Inference Defenses

10. Conclusion

