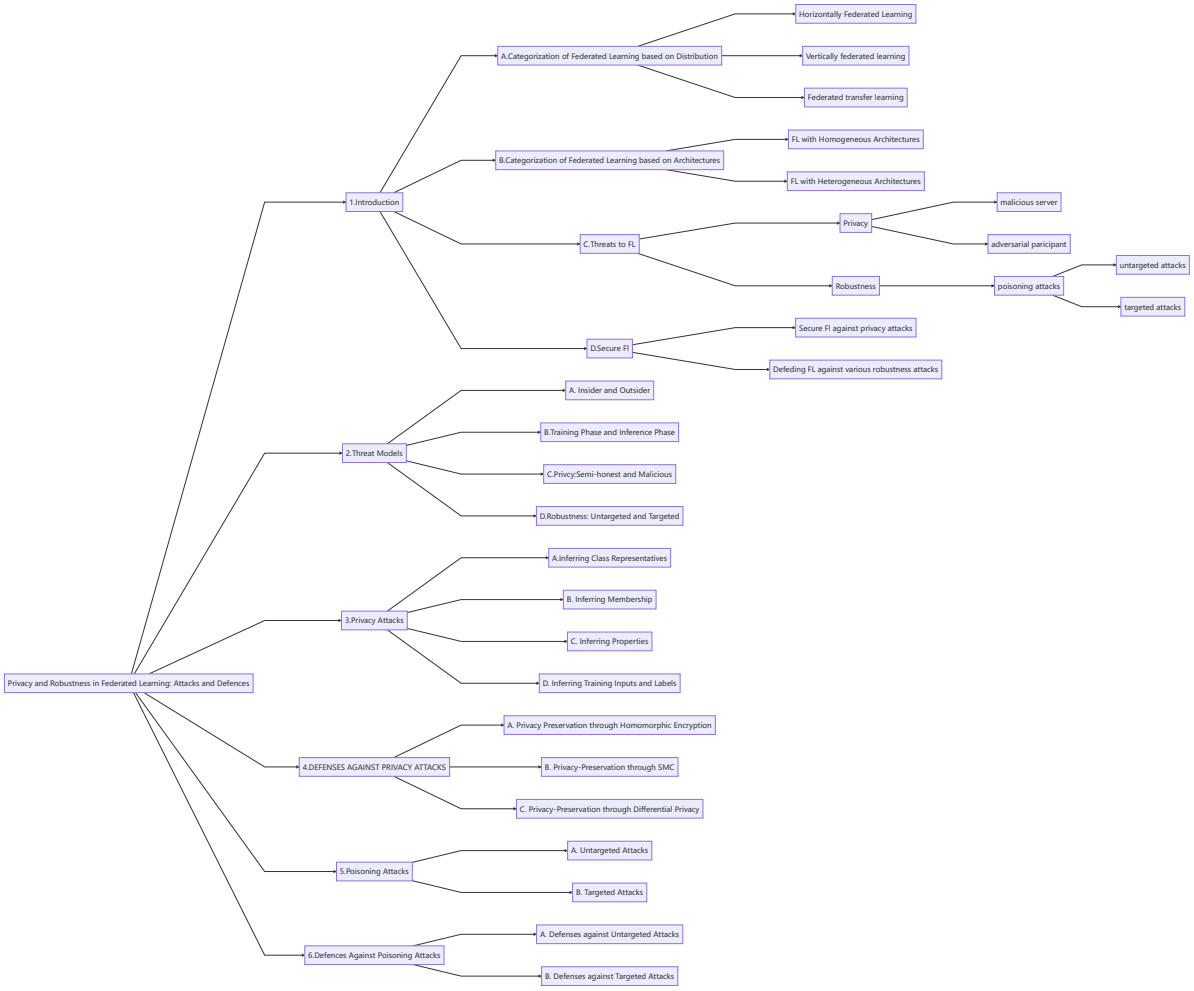
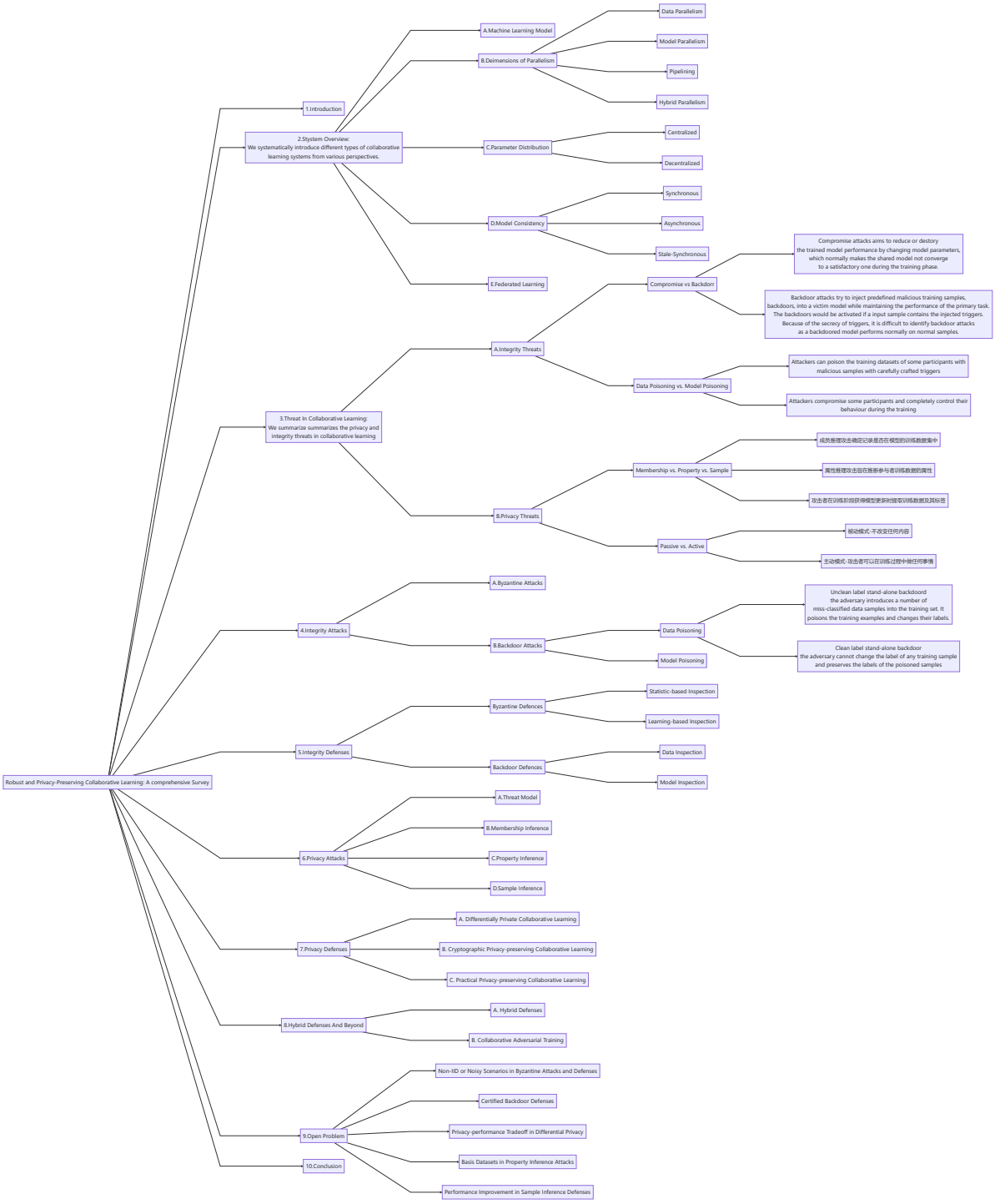
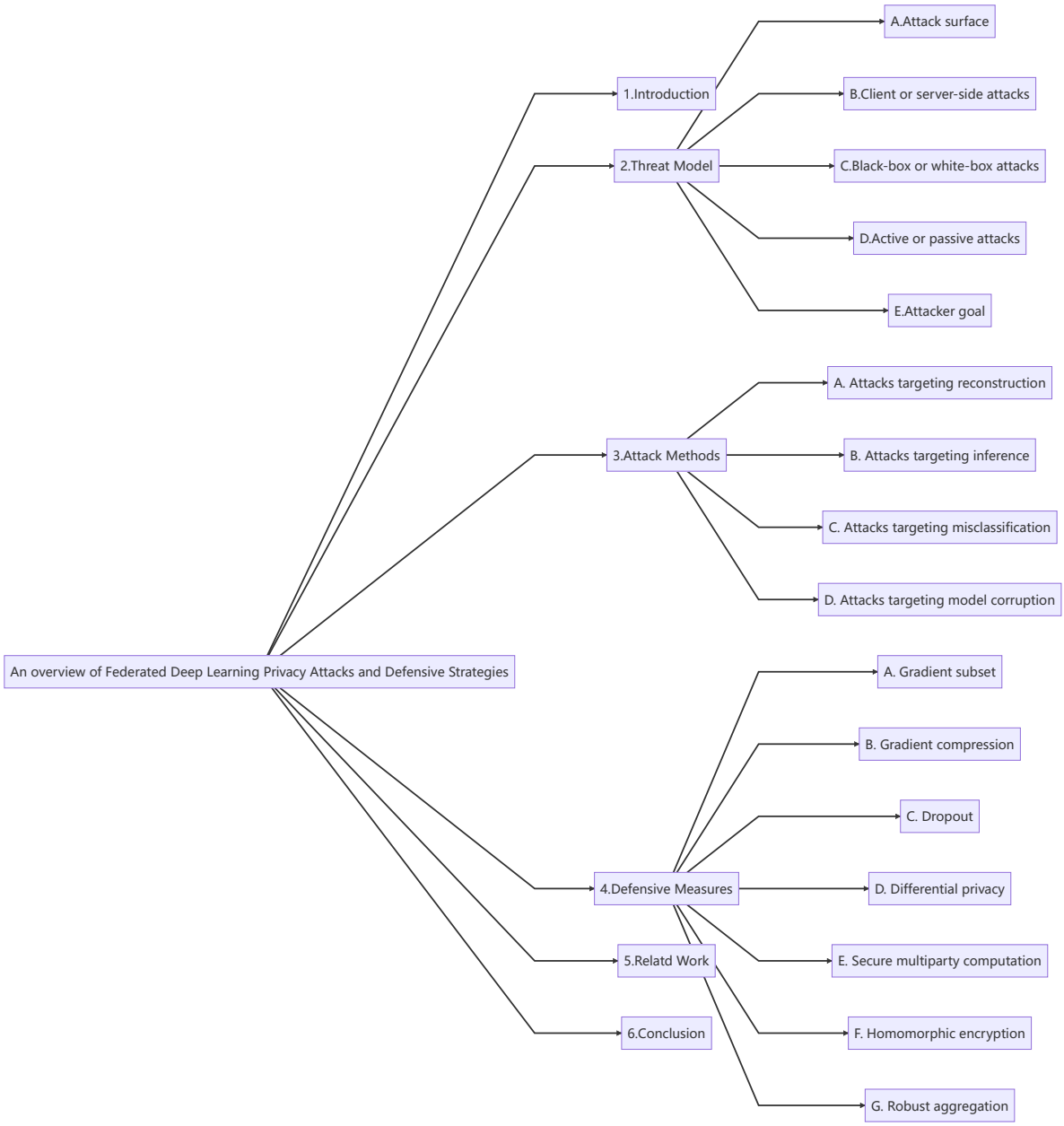
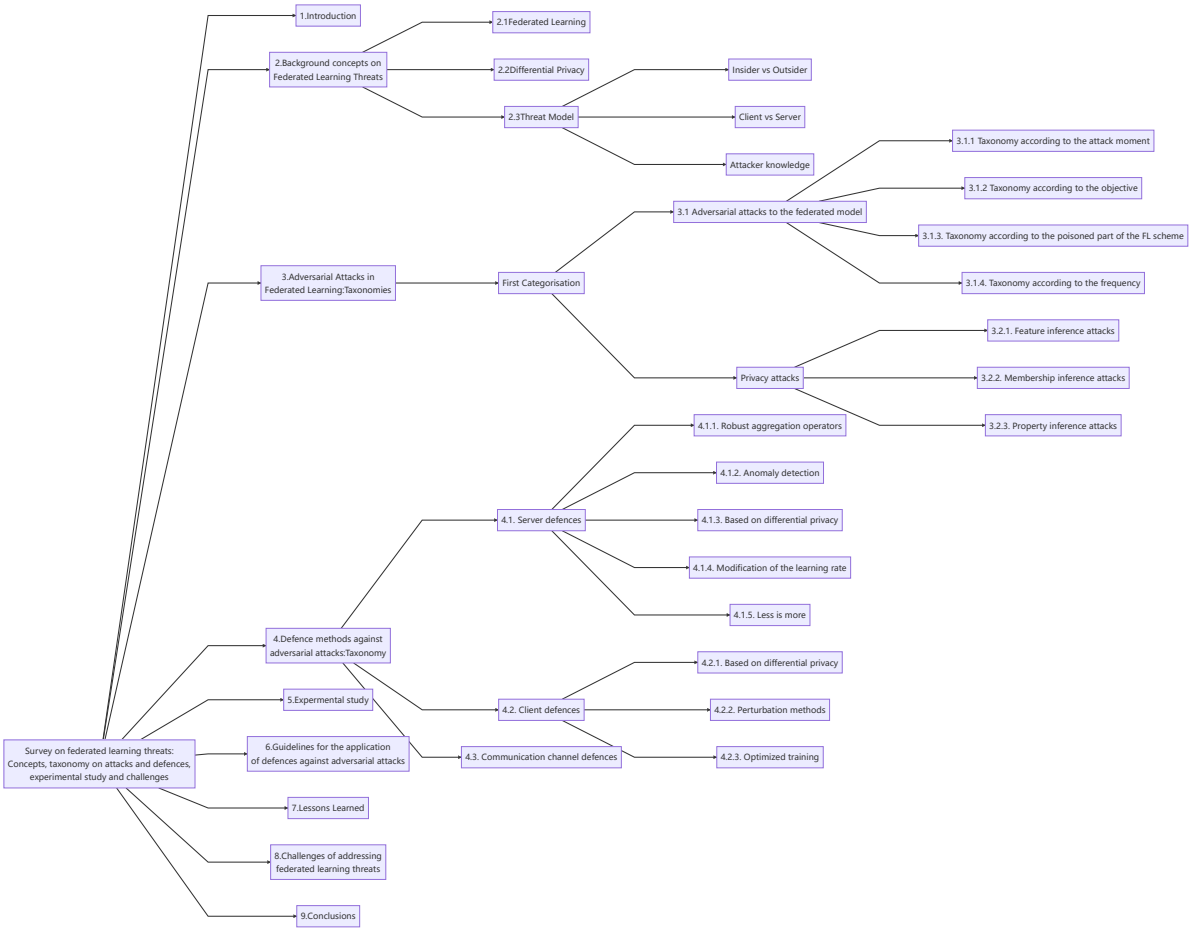
A-Review-Of-Security-Issues-In-Federated-Learning

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
1.Introduction
2.System Overview:
We systematically introduce different types of collaborative learning systems from various perspectives.
A.Machine Learning Model
B.Deimensions of Parallelism
Data Parallelism
Model Parallelism
Pipelining
Hybrid Parallelism
C.Parameter Distribution
Centralized
Decentralized
D.Model Consistency
Synchronous
Asynchronous
Stale-Synchronous
E.Federated Learning
3.Threat In Collaborative Learning:
We summarize summarizes the privacy and integrity threats in collaborative learning
A.Integrity Threats
Compromise vs Backdorr
Compromise attacks aims to reduce or destory 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.
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
攻击者在训练和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 backdoord
the adversary introduces a number of miss-classified data samples into the training set. It poisons the training examples and changes their labels.
Clean label stand-alone backdoor
the adversary cannot change the label of any training sample and preserves the labels of the poisoned samples
Model Poisoning
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


An overview of Federated Deep Learning Privacy Attacks and Defensive Strategies
1.Introduction
2.Threat Model
3.Attack Methods
4.Defensive Measures
5.Relatd Work
6.Conclusion
A.Attack surface
B.Client or server-side attacks
C.Black-box or white-box attacks
D.Active or passive attacks
E.Attacker goal
A. Attacks targeting reconstruction
B. Attacks targeting inference
C. Attacks targeting misclassification
D. Attacks targeting model corruption
A. Gradient subset
B. Gradient compression
C. Dropout
D. Differential privacy
E. Secure multiparty computation
F. Homomorphic encryption
G. Robust aggregation




```
graph LR
a["联邦学习攻击与防御综述"]
a --->a1["1.联邦学习中的攻击类型"]
a --->a2["2.联邦学习中的防御设施"]


a1 ---> a1a["1.1 数据中毒"]
a1 --->a1b["1.2 模型攻击"]
a1 ---> a1c["1.3 推理攻击"]
a1 ---> a1d["1.4 服务器漏洞"]

a2 ---> a2a["2.1 联邦学习通用隐私保护设施"]
a2a ---> a2a1["2.1.1 差分隐私"]
a2a ---> a2a2["2.1.2 同态加密"]
a2a ---> a2a3["2.1.3 秘密共享"]

a2 ---> a2b["2.2 联邦学习针对性防御措施"]
a2b ---> a2b1["2.2.1 防御数据中毒"]
a2b ---> a2b2["2.2.2 防御模型攻击"]
a2b ---> a2b3["2.2.3 防御推理攻击"]
a2b ---> a2b4["2.2.4 防御服务器漏洞"]
```