联邦学习：一种具有隐私保护的分布式学习框架

汇报人：岳高峰
导师：言浬

2023年5月22日

西安交通大学网络空间安全学院
教育部智能网络与网络安全重点实验室

CONTENTS

01

联邦学习简介

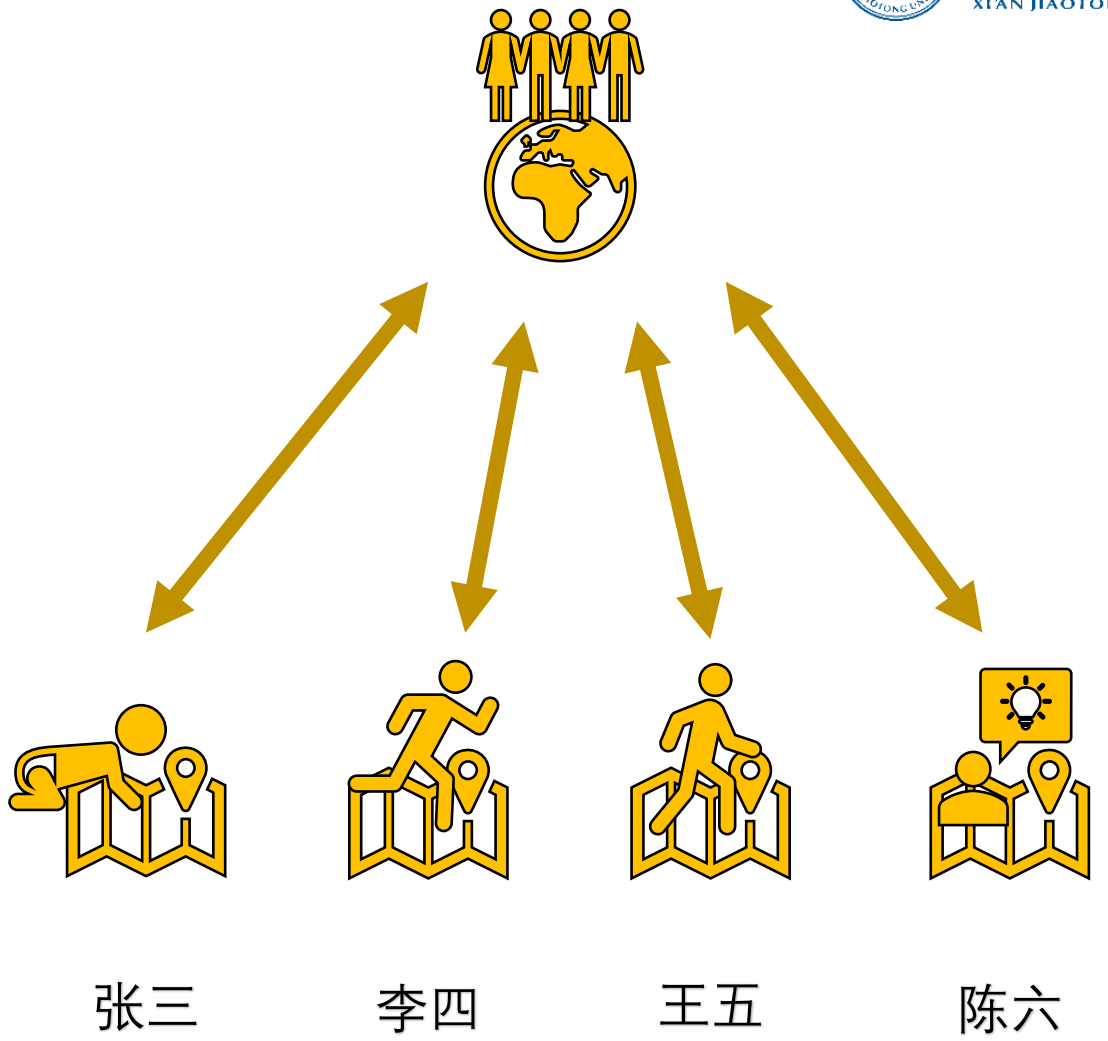
- 产生
- 机理
- 分类

1.联邦学习简介

给定一个假设：

学校为了判断学生的活动区域。
现在需要调取你们每天的轨迹数据！

Can you accept it?



1.联邦学习简介

联邦学习：Federated Learning

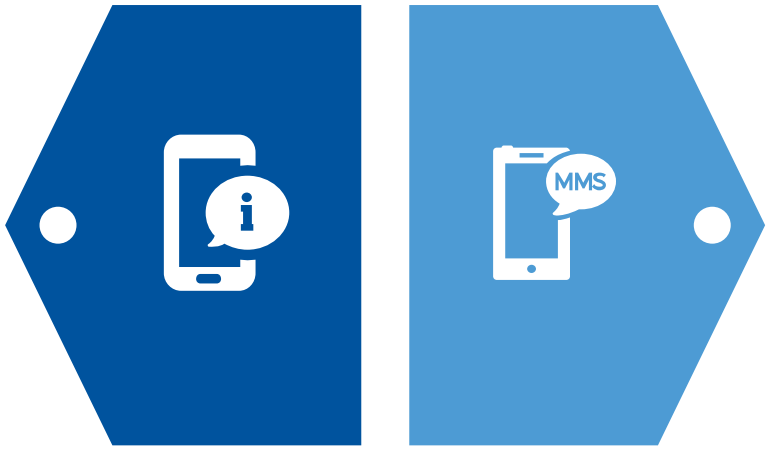
联邦学习（Federated learning， FL）由谷歌于2016年首次提出，旨在解决分布式移动数据集的机器学习的隐私问题。联邦学习将安全技术和机器学习技术相结合，具有保护用户隐私的天然优势，广泛应用在各个领域。

物联网

人体行为监测

智能驾驶辅助

金融保险



智慧城市

交通流量预测

电力能量预测

跨域推荐

1.联邦学习简介

联邦学习机理

1.初始化模型参数

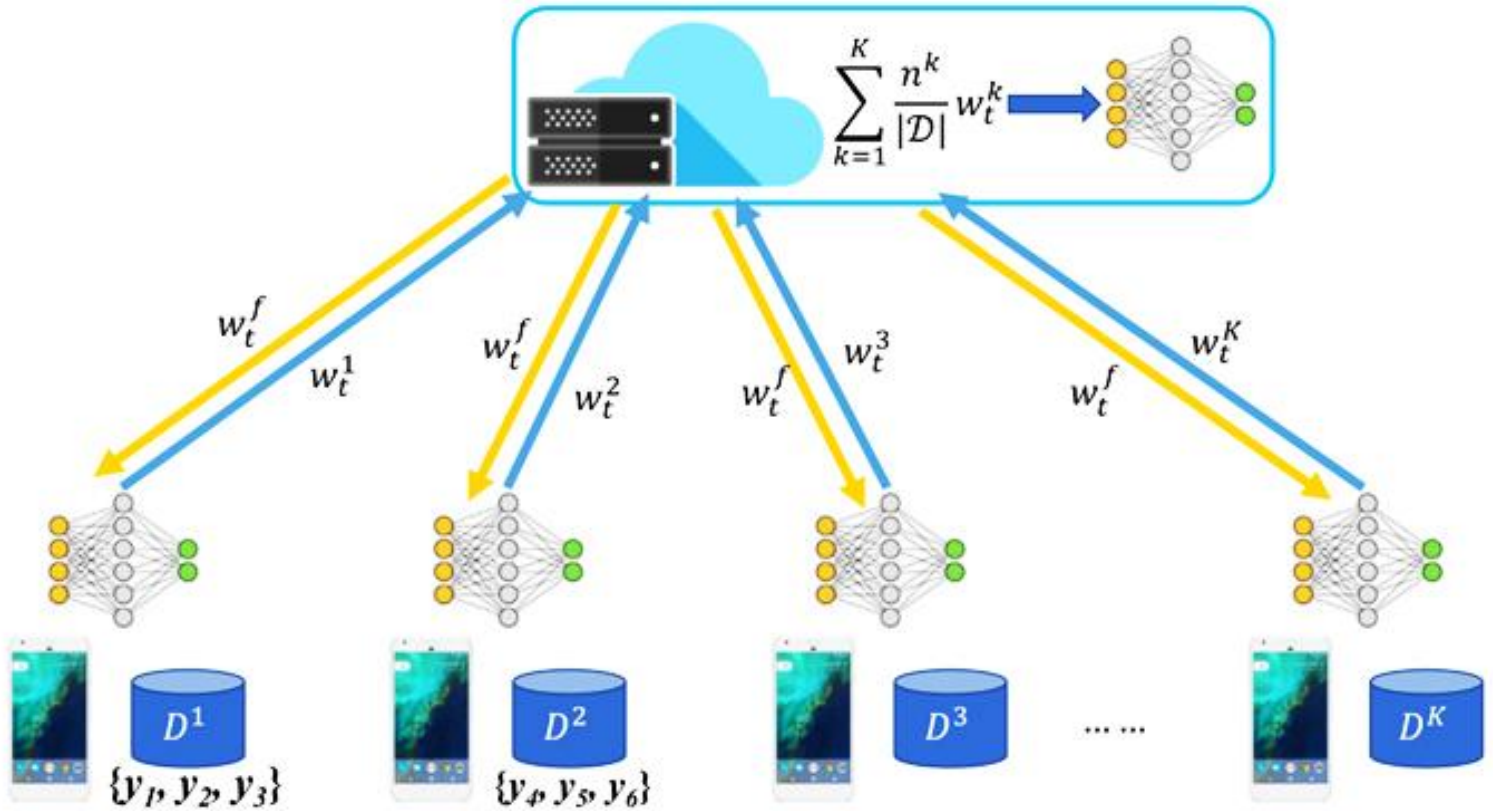
2. 本地模型训练

3.本地发送模型权重

4.联邦服务器聚合

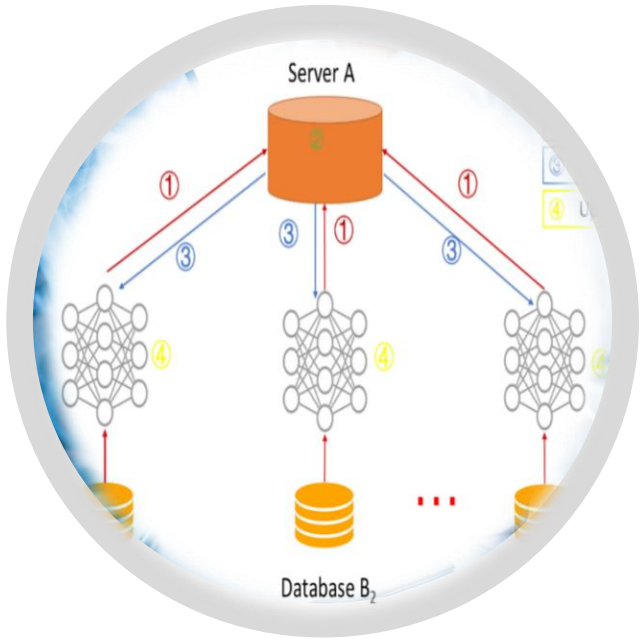
5.反复迭代

6.直到损失函数不变



1.联邦学习简介

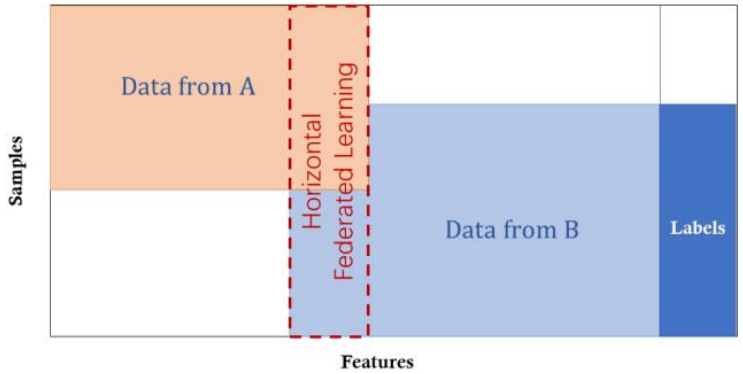
联邦学习分类



1

横向联邦学习

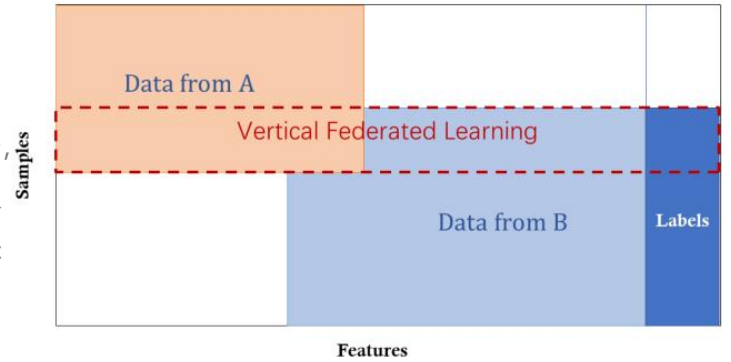
两个地方的银行在各自的地区有不同的用户群体，它们的交集可能非常小，然而，他们的业务很相似，所以特征空间是相同的。



2

纵向联邦学习

同一城市有两个公司，一个是银行，一个是电子商务，他们的用户都是这个区域大部分居民的数据，但是银行记录的是存款、开支情况和信用率，电商记录的是用户浏览和购买的记录，特征空间非常不同



3

联合迁移学习

联邦迁移学习应用在两个数据库不仅样本空间不同，而且特征空间也不同的情境下。假设一个在中国的银行和另一个在美国的电商公司，两者地理条件不同，用户群体交集也很小，另外业务也不同，特征交集也很小。

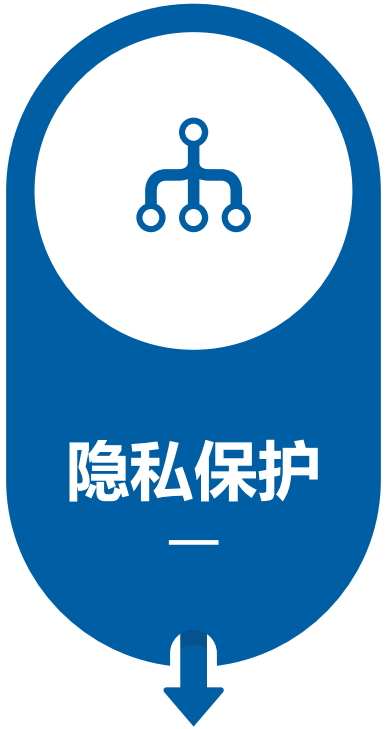
02

联 邦 学 习 优 点

- *隐私*
- *安全*
- *大局*

2.联邦学习优点

西安交通大学 XI'AN JIAOTONG UNIVERSITY



隐私保护

区别于纯分布式机器学习或边缘计算

大大局观

个人利益服从于集体利益，无私奉献

数据孤岛

采用迁移学习解决数据不统一的问题，在AI学习领域是一大突破。

03

联 邦 学 习 缺 点

- *隐私*
- *安全*
- *可信*

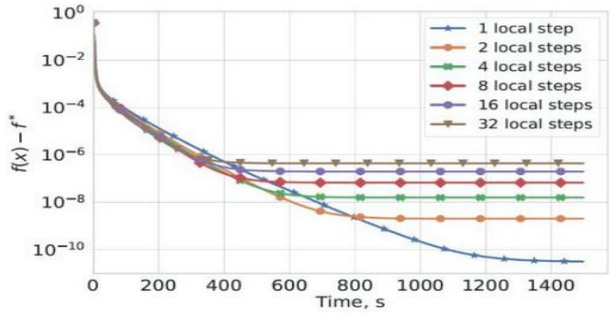
3.联邦学习缺点

隐私保护问题

80%

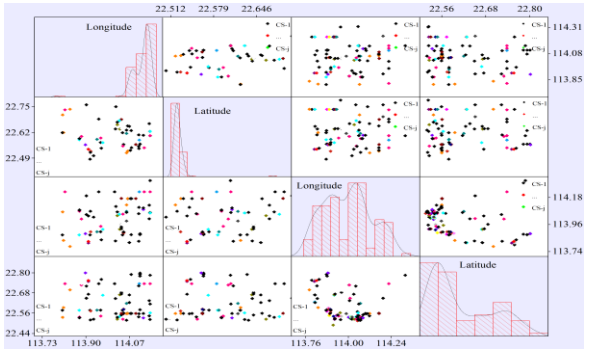
通信算力问题

通信带宽是联邦学习的主要瓶颈，因为大量的设备都将其本地更新发送到中央服务器中



联邦聚合问题

在联邦学习中，存在一个协调方对所有参与方的数据进行安全聚合和运算等。



04

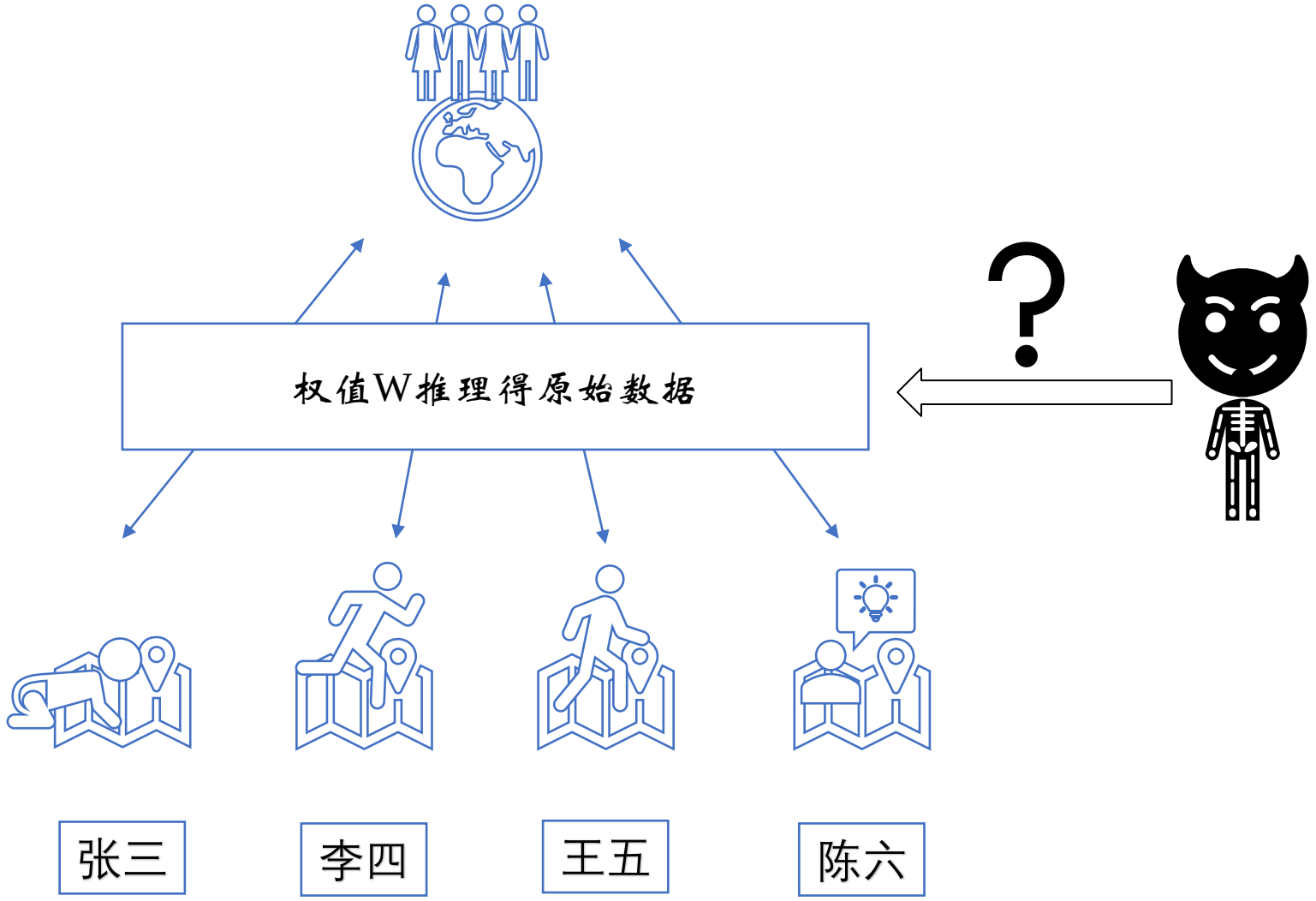
相关研究工作

- *隐私*
- *安全*
- *可信*

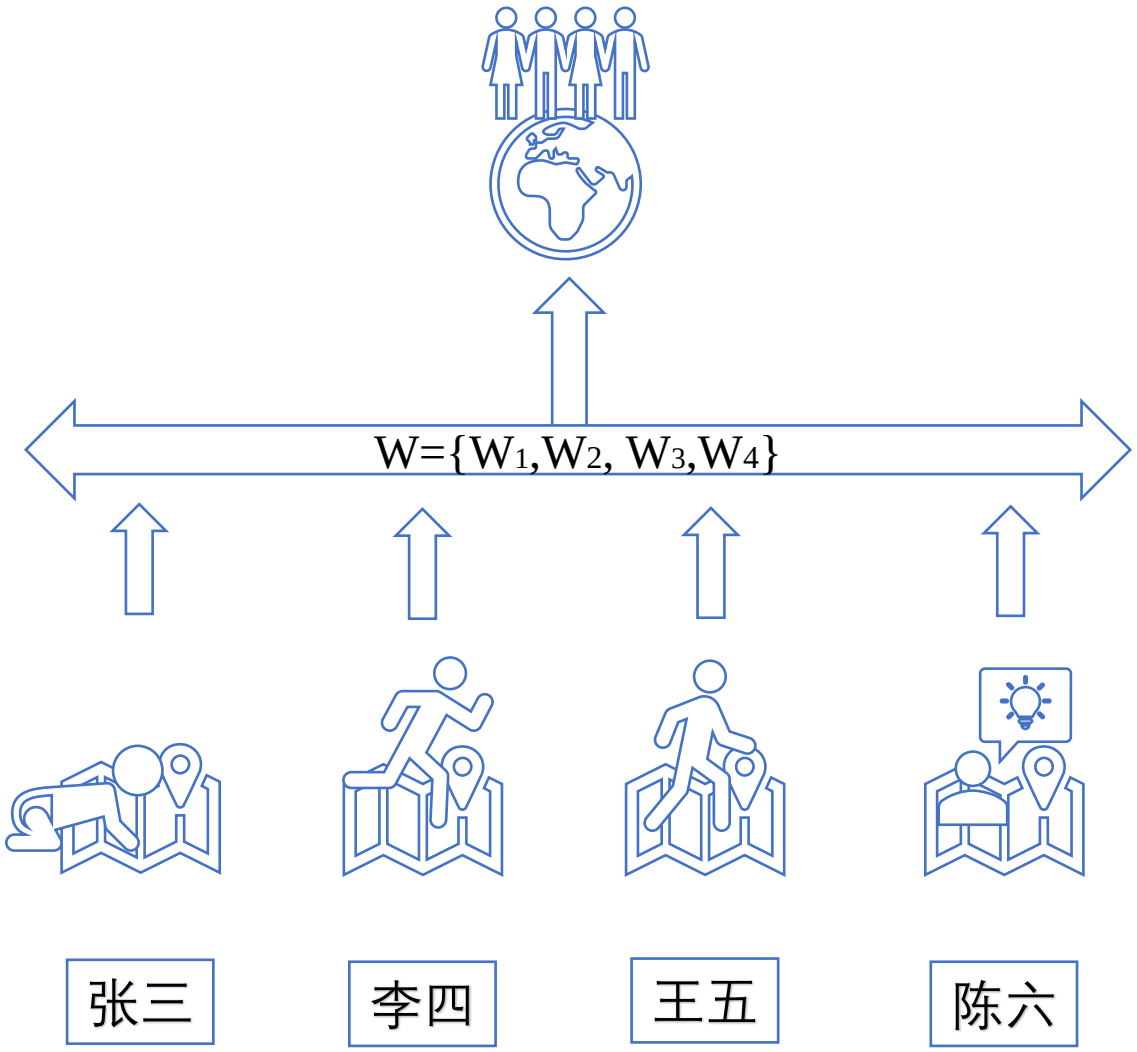
4.相关研究工作

隐私保护问题

80%



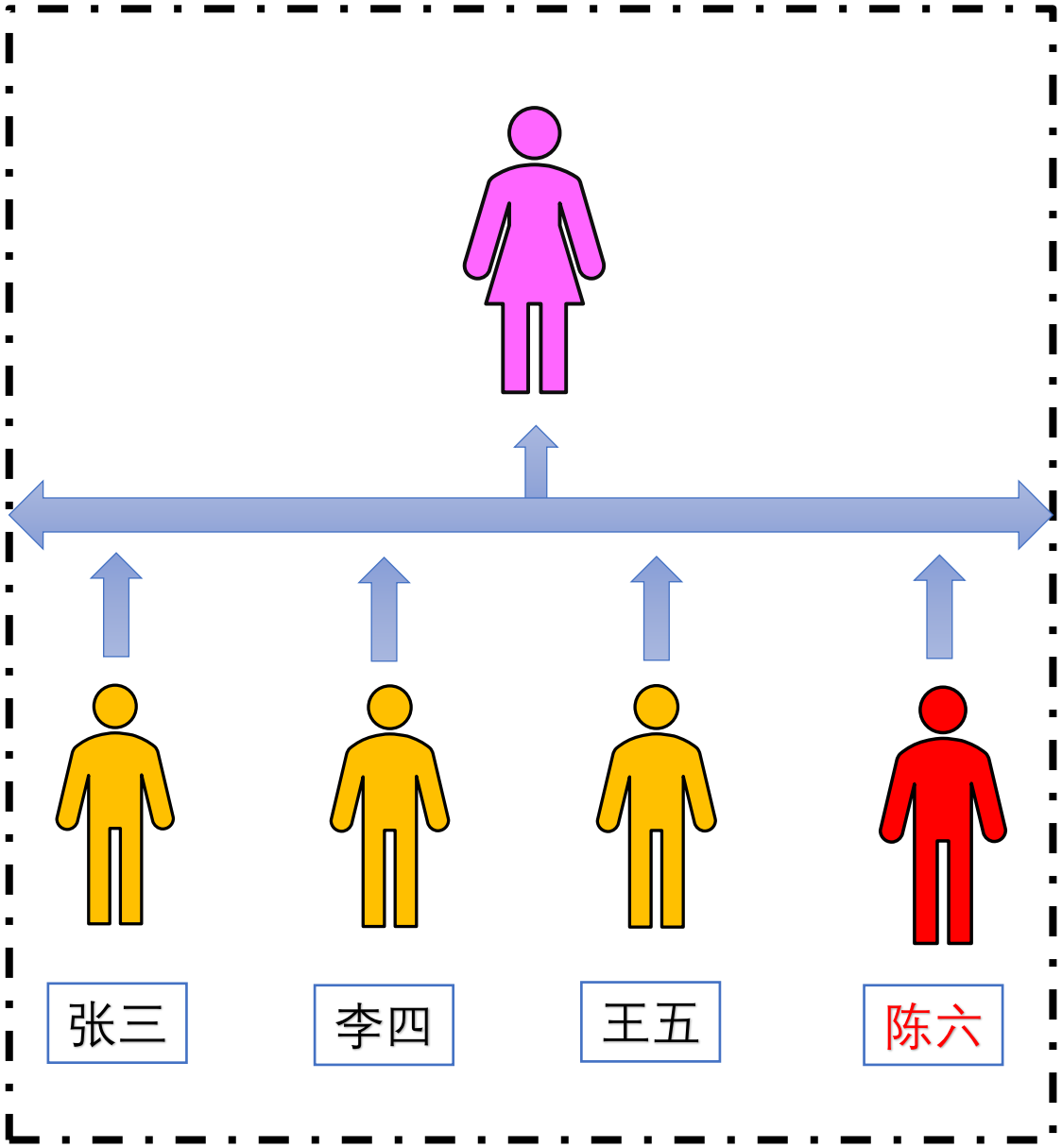
4.相关研究工作-隐私保护问题



硬件解决方案

软件解决方案

4.相关研究工作-隐私保护问题



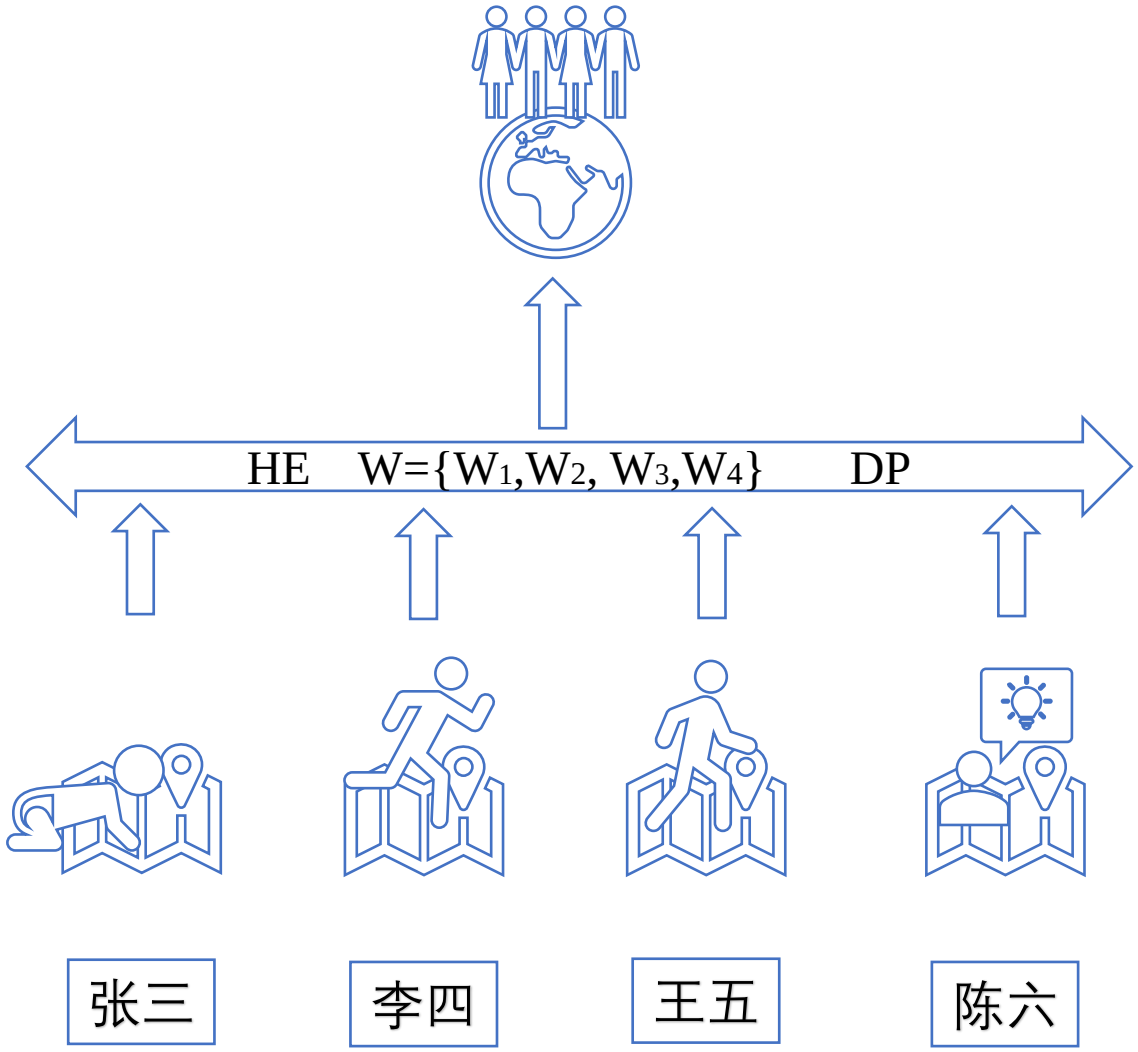
硬件方案：可信计算
把系统关在小黑屋里
（局域网）

Question：半诚实参与方
恶意参与方？

4.相关研究工作-隐私保护问题

隐私保护问题

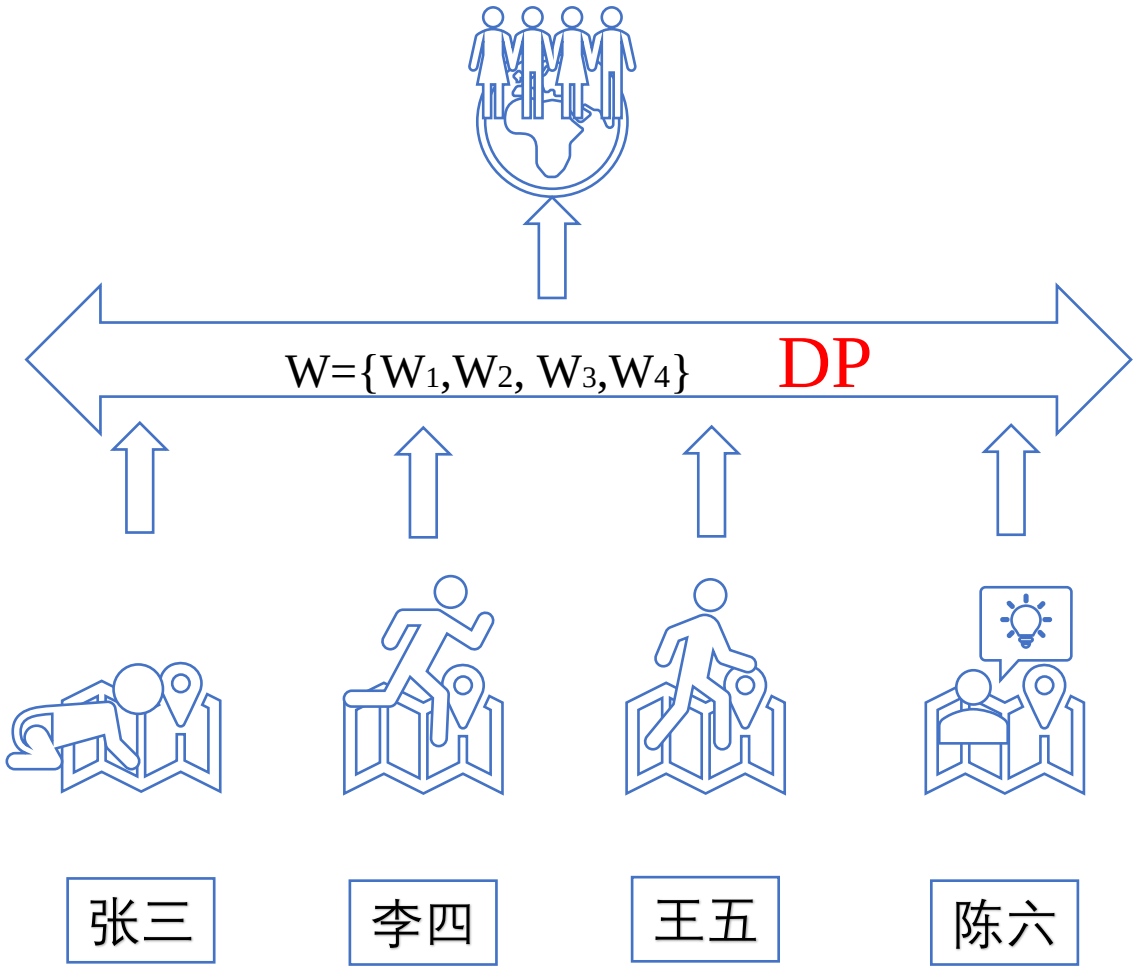
80%



软件解决方案

- 差分隐私
- 加密算法

4.相关研究工作-隐私保护问题



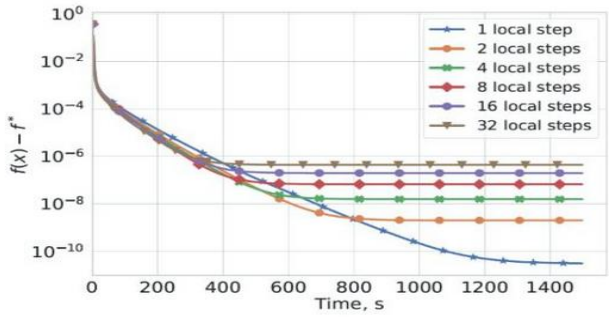
软件解决方案

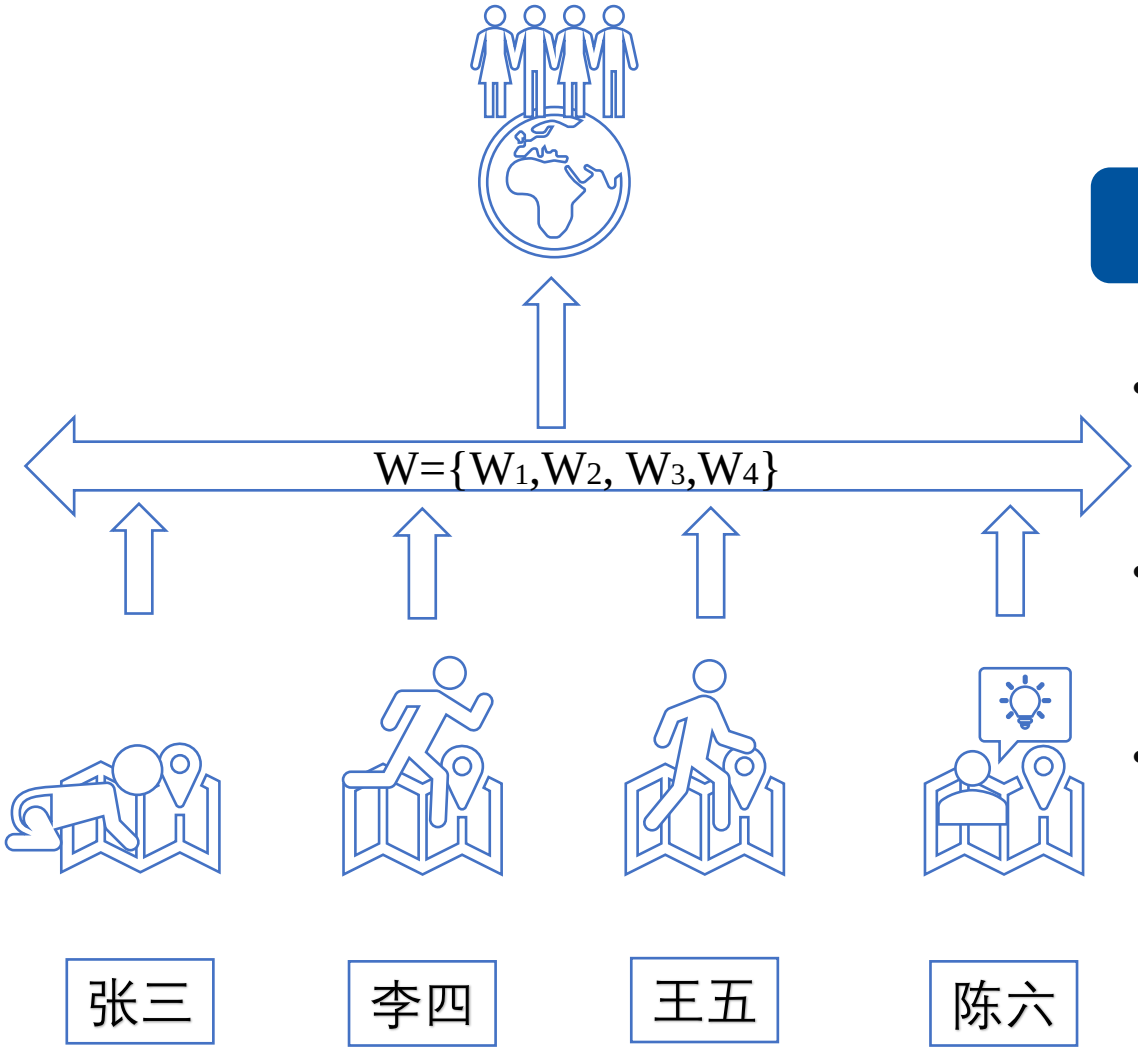
- 差分隐私（differential privacy，DP）
- 基本思路是针对需要保密的数据，加入噪声的处理。该动作虽然有效保护了数据的安全性与隐私性
- 噪声的加入会对计算结果产生一定的影响

4.相关研究工作-通信算力问题

通信算力问题

通信带宽是联邦学习的主要瓶颈，因为大量的设备都将其本地更新发送到中央服务器中

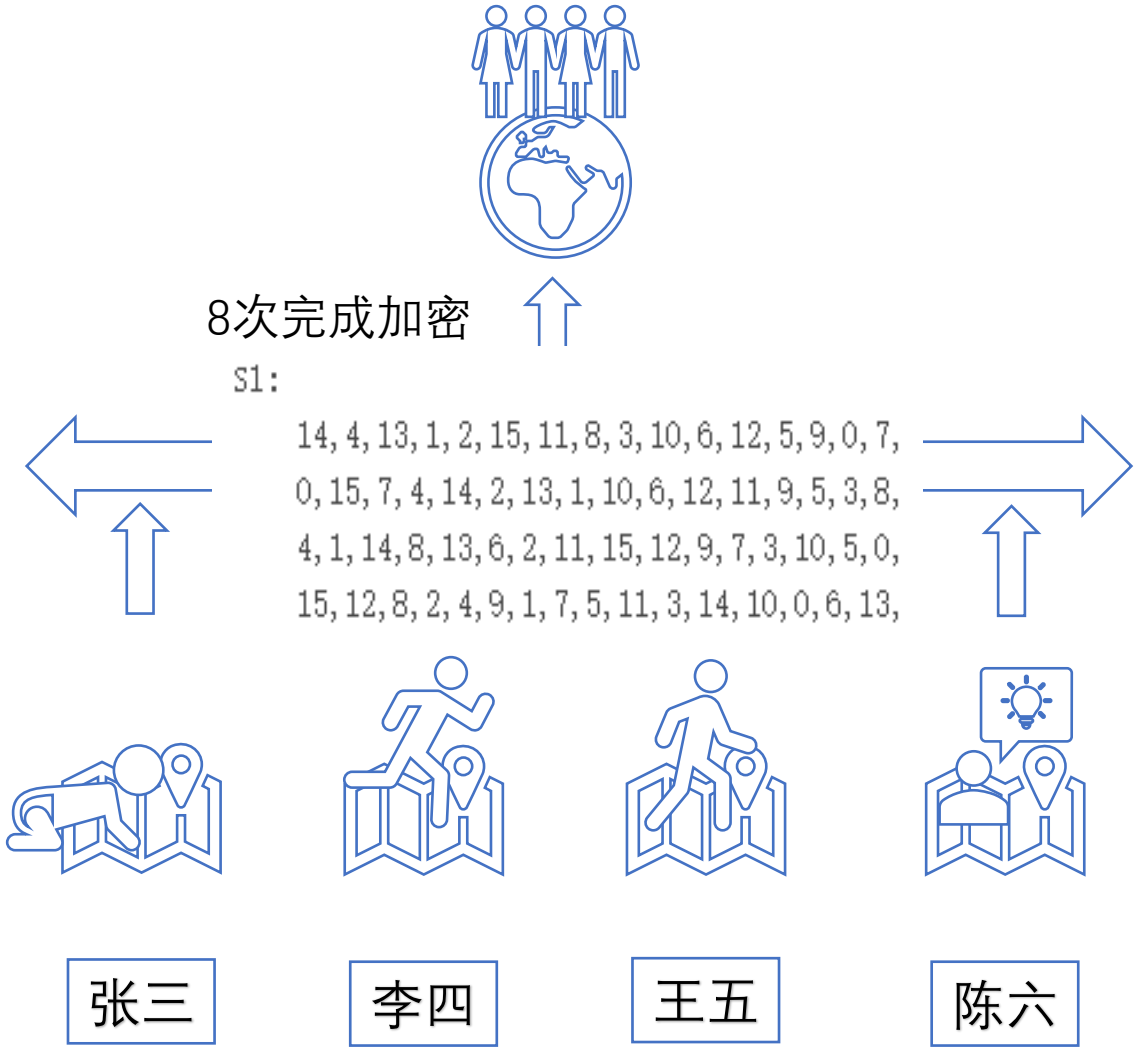




加密算法

- 对称加密算法：DES、3DES、AES 等，
- 非对称算法： RSA、DSA 等
- 散列算法 ： SHA-1、MD5 等。

4.相关研究工作-通信算力问题



加密算法

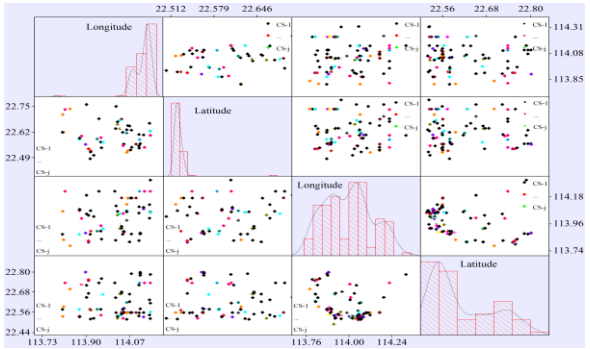
- 对称加密算法：DES、3DES、AES 等，
- 非对称算法： RSA、DSA 等
- 散列算法 ： SHA-1、MD5 等。

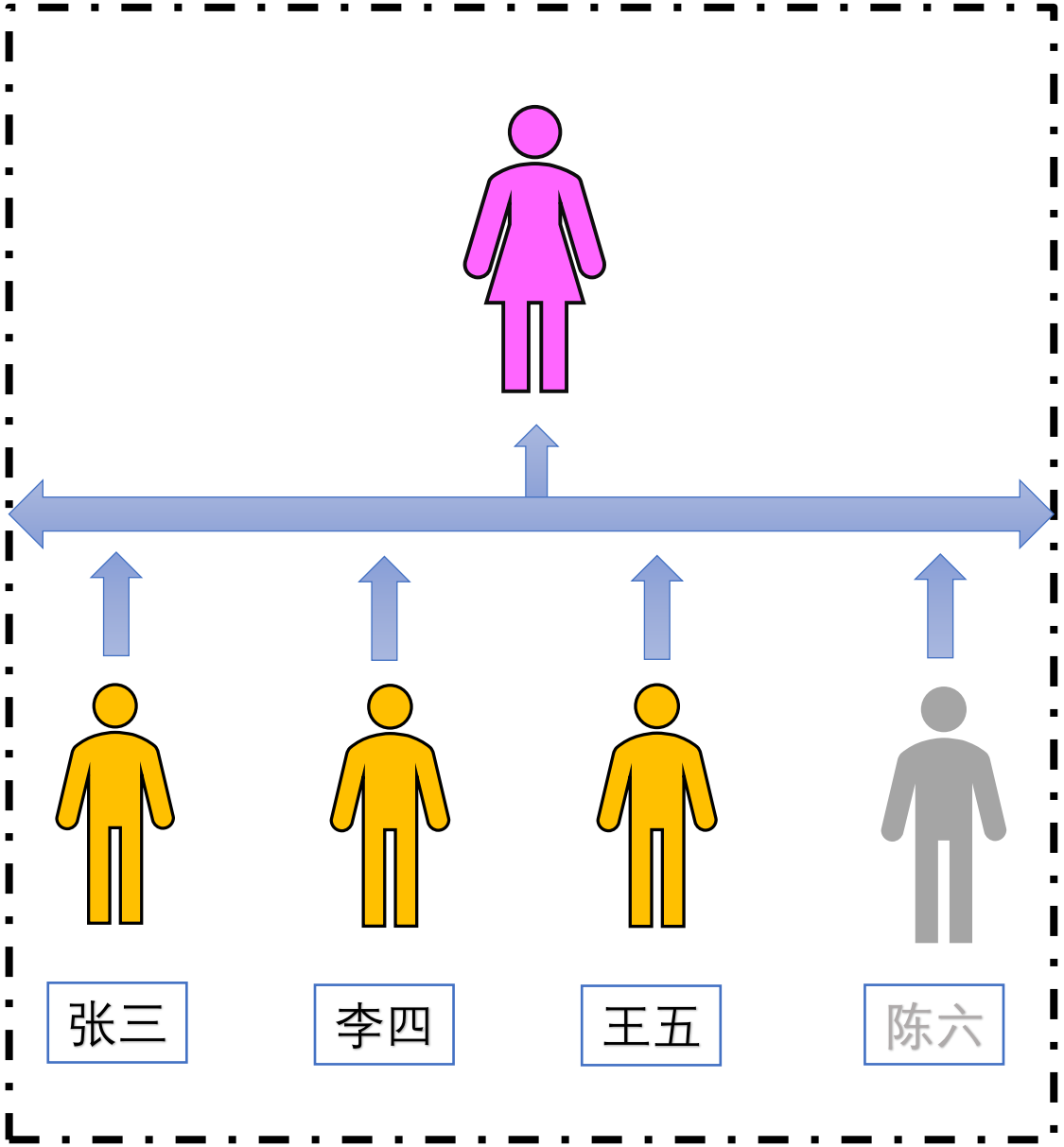
- 虽然可靠，但是通信负载进一步加大。
- 同样存在：半诚实参与方、恶意的参与方

4.相关研究工作-通信算力问题

联邦聚合问题

在联邦学习中，存在一个协调方对所有参与方的数据进行安全聚合和运算等。



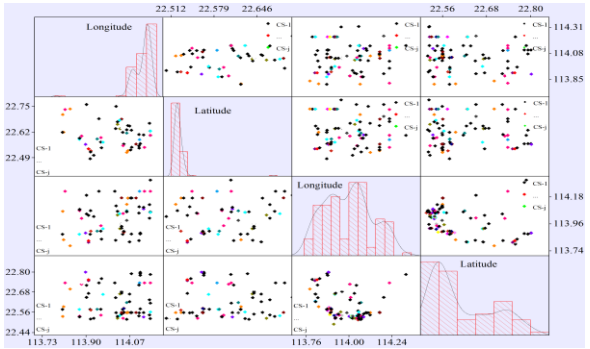


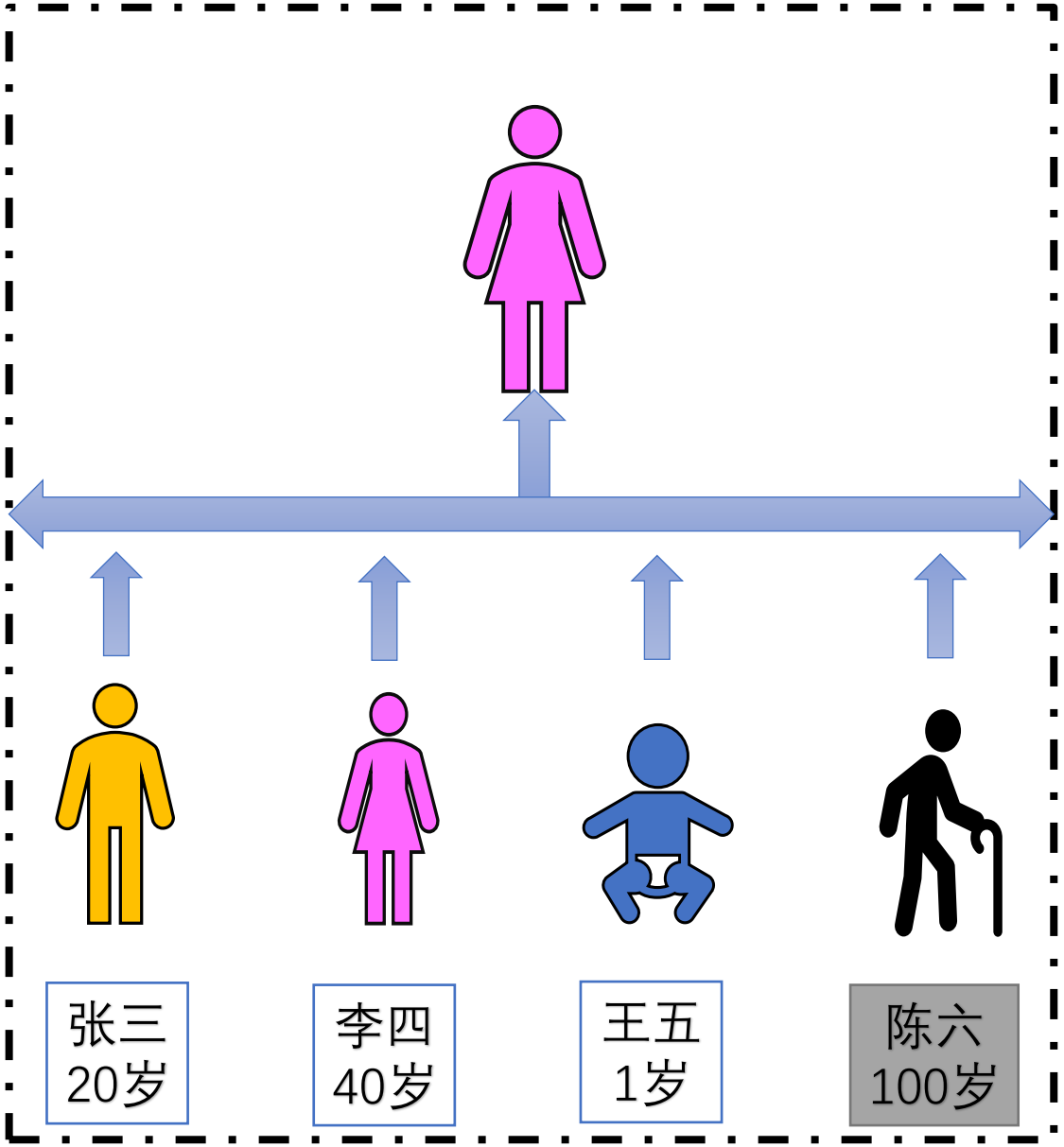
- 如果陈六的网线被拔掉了或者计算机死机了，聚合端应该一直等待吗?

4.相关研究工作-联邦聚合问题

联邦聚合问题

在联邦学习中，存在一个协调方对所有参与方的数据进行安全聚合和运算等。





- 当本地数据呈现异质性（Non-IID）的时候，聚合精度和速度如何保证?

4.相关研究工作-挑战难题

参与方难题

参与方激励难题指的是如何吸引更多的参与方参与到联邦学习中，建立完善的激励机制和分配机制

算力难题

移动设备的算力下仅有部分小运算量的算法如逻辑回归等可在设备端运行

通信难题

如何提高通信信道的质量和容量，成为限制联邦学习发展的难题之一

聚合难题

如何提高通信信道的质量和容量，成为限制联邦学习发展的难题之一

预测难题

在纵向联邦学习中，只有协调方得知的是整个联邦的结构，而参与方得知的是与其数据特征相关的子模型的结构

在如今强调数据权和隐私保护的时代背景下，联邦学习具有巨大的前景

05

个 人 工 作 总 结

- *安全*
- *高效*
- *可用*

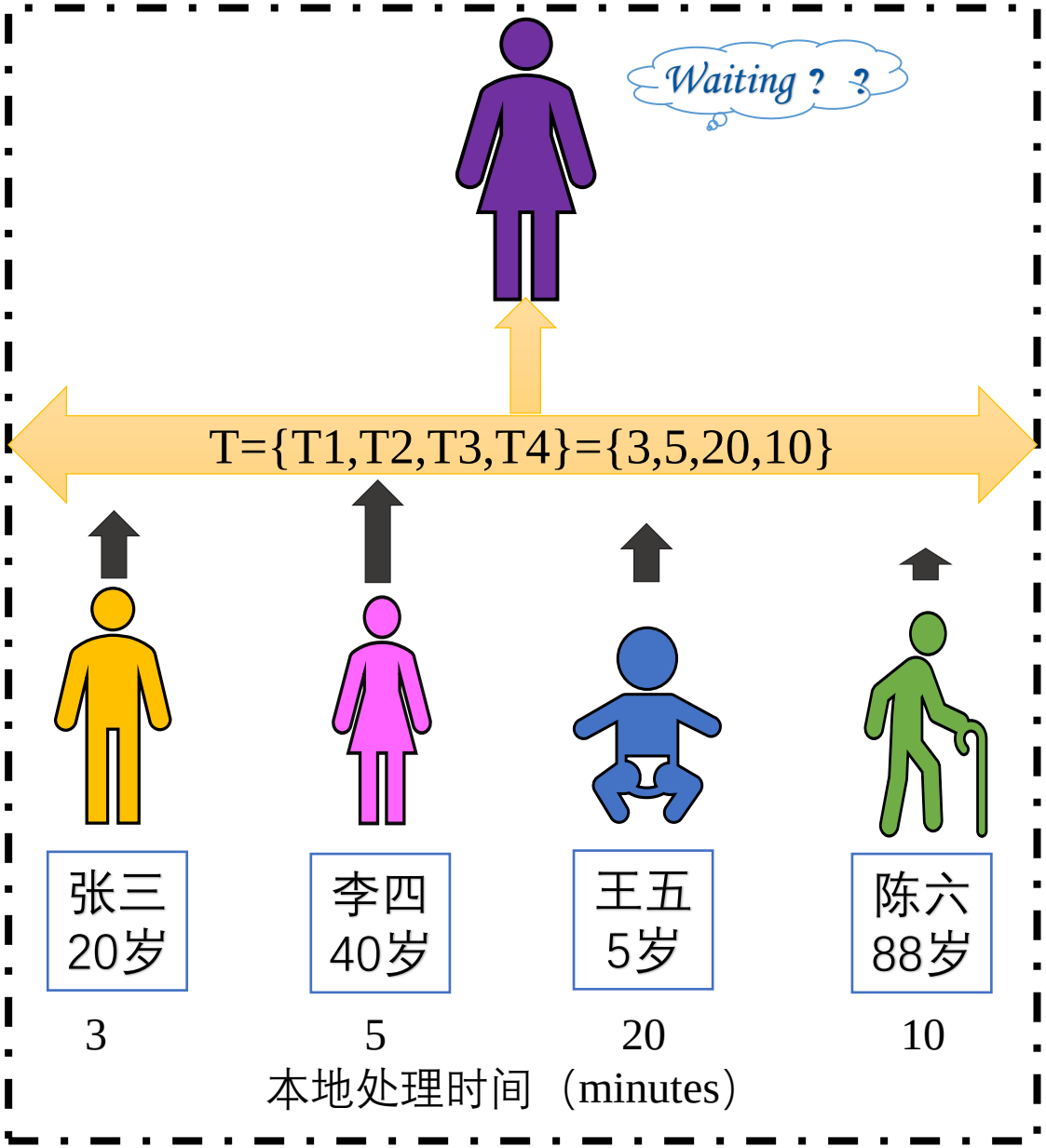
个人工作简介1

异质异步问题：

本地客户端处理性能不同，时间消耗不同，
服务器应该等待吗？
时间不同步导致训练时间长，
数据异质导致模型精度不足，
提出一种异质异步联邦学习框架。

Heterogenous Asynchronous Federated Learning

AHFL



个人工作简介2

西安交通大学 XI'AN JIAOTONG UNIVERSITY

隐私保护问题

80%

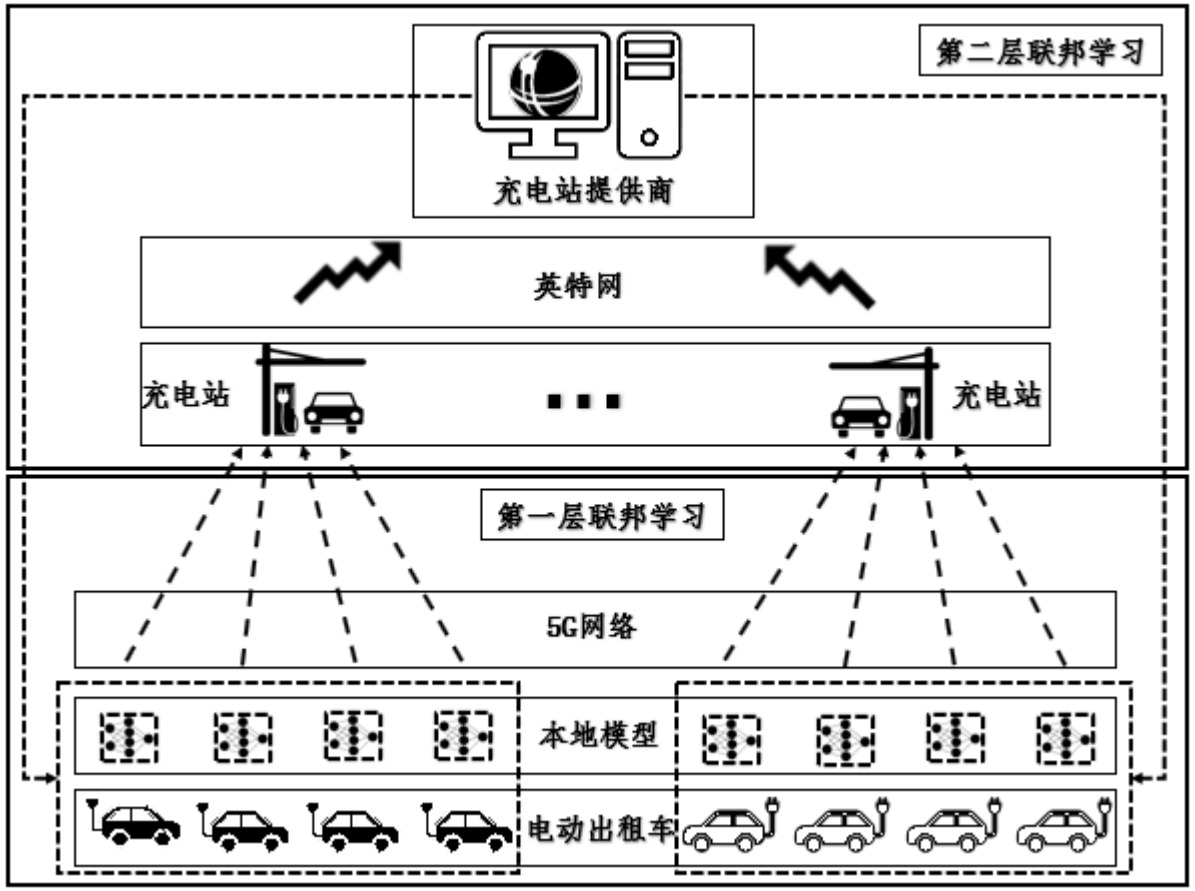
Q2：有没有一种方法？

既保护用户隐私
又能加速聚合
还能解决不诚实参与方问题？

基于梯度修正的动态随机森林聚合法
问题是：不能100%完全隐私保护（90%）

个人工作简介3

Q3：很多情况下，简单分布式模型并不能系统地反应和解决问题，实际都是金字塔结构。



西安交通大学
XI'AN JIAOTONG UNIVERSITY

谢谢大家

汇报人：岳高峰

西安交通大学网络空间安全学院
教育部智能网络与网络安全重点实验室

2023年5月

西安交通大学
XI'AN JIAOTONG UNIVERSITY

欢迎提问

汇报人：岳高峰

西安交通大学网络空间安全学院
教育部智能网络与网络安全重点实验室

2023年5月