2023年无线网络安全技术上机计划

1. **上机分组**：3-4人一组（自行组队），不足3人也可以；
2. **上机任务**：针对某种数据（例如合成数据，轨迹数据、医疗数据、社交关系数据等）完成一份完整的数据发布匿名保护方案。
3. **背景知识**：未经匿名保护的数据在发布出去的时候，容易受到各种隐私攻击的威胁，例如Record Linkage Attack等，因此在发布数据之前，需要对原始数据进行匿名处理，但是不能采取数据加密方法。如果对数据进行加密之后，仅持有密钥的人才能访问数据，违反了公开发布数据的原则。本项目目的在于通过调研，采用现有的或者自定义隐私攻击，保证发布数据中敏感数据隐私同时，提供一定的可用性。

参考资料，包括但不限于：

第1类：隐私程度估量与数据发布保护机制

（1）2009-Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis-Privacy Accountant Concept

（2）2010-Closeness\_A\_New\_Privacy\_Measure\_for\_Data\_Publishing

（3）2014-RAPPOR\_Randomized Aggregatable Privacy-Preserving Ordinal Response

（4）2015-Privacy-Preserving Compressive Sensing for Trajectory Recovery in Mobile Social Networks

（5）2017-Fisher Information as a Measure of Privacy-Preserving Privacy of Households With Smart Meters Using Batteries（较难）

（6）2018-Privacy\_Characterization\_and\_Quantification\_in\_Data\_Publishing（较难）

（7）2019-Optimal Utility-Privacy Trade-off with Total Variation Distance as a Privacy Measure（较难）

（8）2020-Information theoretic-based privacy risk evaluation for data anonymization（容易）

（9）2022-Bridging unlinkability and data utility\_Privacy preserving data publication（容易）

（10）2022-Data Privacy and Utility Trade-Off Based on Mutual Information Neural Estimator（容易）

**第1类文献中，文献（5）-（7）包含隐私程度的估量方法和基于此估量方法隐私保护方案，如果时间不够，可以实现隐私估量方法，隐私保护机制做出了适当加分。**

**第1类中的其余文献主要以隐私保护方案为主，隐私估量为辅助环节，较为容易。**

第2类：生成对抗网络及其隐私保护

（11）2014-Generative Adversarial Nets-NIPS-Paper

（12）2016-ICLR-DCGAN

（12）2016-Coupled Generative Adversarial Networks

（14）2018-RDCGAN: Unsupervised Representation Learning with Regularized Deep Convolutional Generative Adversarial Networks

（15）2019-Generalization in Generative Adversarial Networks: A Novel Perspective from Privacy Protection

（16）2019-PPGAN\_Privacy-Preserving\_Generative\_Adversarial\_Network

（17）2019-Pate-gan: generating synthetic data with differential privacy guarantees

（18）2019-Differentially\_Private\_Mixture\_of\_Generative\_Neural\_Networks

（19）2019-Differential Inference Testing: A Practical Approach to Evaluate Sanitizations of Datasets

（20）2021-P3GM：Private High-Dimensional Data Release via Privacy Preserving Phased Generative Model

**文献（11）-（14）主要侧重于生成对抗网络生成的合成数据；**

**文献（15）-（20）关注生成对抗网络和隐私保护的结合，其中（17）-（20）在生成对抗网络中融入隐私保护的方法，可以在确保生成对抗网络能正常运行的情况下，加入隐私保护机制，隐私保护机制做出了会适当加分。**

4、**作业说明**：

1. 源代码，需要加注释说明
2. 预处理后的数据及其对应源码

备注：

1. 允许针对已经调研到的别人的算法进行精简，即别人能够抵御两类攻击，你可以只抵御一种，但是需要保证原理上的正确性；
2. 允许对参考资料中或其他隐私安全的文献算法进行概述和复现，但是需要注明出处，但尽量避免与其他组的同学重复；
3. 可以采用机器学习、深度学习、分布学习等方法进行算法修改，达到同样目的。
4. 为了减轻大家工作量，不需要交文档，如果组长没有对组员工作进行特殊说明，按照小组成果整体给分数。
5. 如果对上机内容感兴趣，可以在小学期或者本科毕设继续做，需要提前联系老师。