Master Thesis Projects in Cyber Privacy

Abbreviations

GDPR	General Data Protection Regulation
PIR	private information retrieval
PC	private computation
DIS	distributed information system
QDIS	quantum DIS
IT	information theory
CT	coding theory
IoT	Internet of things
ML	machine learning
FL	federated learning

1 User Privacy in Emerging Technologies

The right to privacy in today's modern age of information is of vital importance. Society is starting to realize that the privacy and integrity of data stored in public and private databases needs to be protected. One important example of this increasing awareness of the issue is the implementation of the General Data Protection Regulation (GDPR) by the European Union in 2018. This project addresses *fundamental* research related to ensuring privacy in modern information systems with an emphasis on distributed technologies which are a key ingredient in the Internet of Things (IoT), edge caching and computing in fifth-generation (5G) wireless networks, and machine learning (ML).

IoT

The IoT promises to improve individuals' quality of life, from health to safety. However, the massive scale of data sharing in the IoT poses significant privacy challenges, which, if not addressed, may hamper its ubiquitous adoption by privacy-concerned individuals. Inspired by the methods of private information retrieval (PIR) that guarantee strong privacy for the users, we are interested in deriving new privacy-preserving schemes for the IoT scenario.

Edge Caching/Computation

In order to reduce energy consumption and perform computationally-intensive tasks, offloading client computations has been proposed as an efficient solution. However, for delay-sensitive applications, e.g., autonomous driving, offloading computations closer to the clients, e.g., to the *wireless edge* instead of the cloud, is essential in order to be able to meet strict computation deadlines. By the wireless edge we refer to the end nodes of a network, e.g., the small base stations of a 5G network. In a similar manner, in order to reduce network traffic congestion during peak hours and the overall delay for content delivery in general, popular content can be stored at local cache memories without knowledge of later demands. Content can be either cached in

the clients' local cache memories or at the wireless edge. With the ever-increasing proliferation of smartphones and the rapid deployment of IoT devices, distributed caching and computing at the edge have been investigated extensively in recent years. However, less work has been devoted to the study of privacy issues. By combining contemporary cryptographic primitives like PIR, private computation (PC), differential privacy, or secret sharing with modern wireless communications technologies, like MIMO and in particular beamforming in order to exploit multicasting opportunities over the wireless channel, we believe that current solutions can be significantly enhanced.

Distributed ML / Federated Learning (FL)

In ML, since model training may be done on sensitive personal data, such as our health records or financial transactions, privacy-preserving mechanisms have been studied actively over the last decade, albeit to a lesser degree from the perspective of multiple parties. This knowledge gap should be closed as learning in a distributed manner will significantly improve performance and can offer a more effective solution. By allowing local training data to stay local, a distributed ML paradigm referred to as FL was proposed recently to allow for parallelization and to guarantee some level of user privacy. The idea is to train models on local datasets and aggregate these models into a single, stronger model. In FL nodes periodically send their local models to a coordinator that aggregates them and redistributes the aggregation back to continue training with it. However, FL poses several serious challenges, such as, e.g., privacy leakage to strong adversaries both in the upload and download and also high communication cost among entities. On the other hand, deep neural networks have shown remarkably effective for many ML tasks. An alternative fascinating topic in ML is to find good solutions for enhancing privacy-preserving technologies through deep learning algorithms. In particular, we are interested in implementing practical privacy-preserving schemes obtained by deep learning in real databases, e.g., servers owned by Amazon. Having experience or a strong interest in deep learning and probability theory is an advantage.

Quantum Distributed Information Systems (QDISs)

Quantum computing has received significant theoretical attention in recent years. In the future, it is likely that data can be stored across quantum computers in a distributed manner. In order to design privacy-preserving mechanisms for QDISs, a fundamentally new understanding and implementation based on quantum information theory (IT) and coding theory (CT) will be required. Recently, another focus area of our group is the design of cost-efficient PIR or PC protocols for QDISs. We are looking for students who are willing to learn the relevant theories for quantum privacy-preserving technologies.

2 What Types of Theses Can We Supervise?

We can supervise theses both aiming at theoretical aspects (e.g., discovering and proving theoretical facts, algorithms and protocols, analyzing their theoretical complexity, etc.) and practical implementation of known theoretical results in realistic practical environments. The topics above are intentionally formulated in rather general form. The particular details will be discussed and agreed with a potential student personally, based on his/her background and interest.

NB! If you are unsure whether a topic really fits your interests, you are more than welcome to contact us for more details and discussion.

Theoretical Theses

In this type of theses, we would like to improve existing theoretical results, which may include—but not limited to—new optimal and/or heuristic algorithms, theoretical bounds on possible values of involved variables, and so on and so forth. A decent mathematical background is required. We will use tools from linear algebra, probability and number theory, as well as IT. In order to perform calculations to illustrate theoretical findings, some programming skills is a plus but not strictly required.

Implementation Theses

In this type of theses, we will concentrate on existing theoretical approaches and known theoretical schemes and the goal of the project will be to create a working implementation in an environment close to real-world applications. A strong mathematical background is not required. However, some basic understanding of linear algebra and IT is needed in order to grasp the existing theoretical results. We would be glad to provide any kind of support and tutoring. Since a thesis would be of implementation nature, decent programming skills are essential.

3 About Us

The Master thesis will be supervised by Dr. Yauhen Yakimenka (postdoctoral fellow), Dr. Hsuan-Yin Lin (research scientist), and Dr. Eirik Rosnes (leader of the Information Theory Section), with different degree of involvement based on the particular project. We have a solid experience in IT and CT, as well as many related areas. Currently, our research focus includes various topics in PIR, PC, and privacy-preserving technologies in general. We mostly consider the research problems from the point of view of IT.

Contact Details

- Dr. Yauhen Yakimenka, `yauhen@simula.no`
- Dr. Hsuan-Yin Lin, `lin@simula.no`
  web: `http://hsuanyin-lin.com/main.html`
- Dr. Eirik Rosnes, `eirikrosnes@simula.no`
  web: `https://sites.google.com/a/simula.no/eirik-rosnes/home`