Master Thesis Projects in Privacy and Quantum

Abbreviations

PIR	private information retrieval
PC	private computation
IT	information theory
CT	coding theory
ML	machine learning
FL	federated learning
DL	decentralized learning
NISQ	noisy intermediate scale quantum
QDIS	quantum distributed information systems

1 User Privacy and Security for Emerging Technologies

Nowadays, for cost-effective reasons, information is stored in a distributed fashion. Examples are cloud storage, data centers, etc. Due to past research efforts that keep refining the design of flexible, reliable, and inexpensive distributed storage systems (DSSs), we can ubiquitously enjoy a very high level of data accessibility today. Moreover, intensive work on the security of DSSs allows for the data stored in the cloud to be protected against a variety of possible attacks. Unfortunately, less work has been devoted to the study of privacy issues at the user level.

Recently, society is starting to realize that the privacy and integrity of data stored in public and private databases need to be protected. In this project, we aim to develop efficient and practical privacy-preserving and secure protocols. The primary goal is to address the **design of cost-efficient practical information-theoretic protocols for providing privacy and security in distributed information systems**, including the blockchain and distributed machine learning (ML).

Blockchain

One of the main issues of current blockchain systems is that each network node must store the *complete information* and verify the complete chain. This prevents scalability and poses a main bottleneck but also provides high security (in the sense of tolerating many adversarial nodes in the system). Thus, *sharding* (see, e.g., Ethereum-Wiki), which allows to split the blockchain up and to handle much smaller parts in a distributed fashion, has been proposed. However, due to this splitting, the security level of the system is reduced (far fewer adversary nodes than before can be allowed). To address this shortcoming, the very recent concept of *coded sharding* was proposed, and it raises a new challenge of designing blockchain systems that are simultaneously able to verify a huge number of blocks (for scalability) and tolerate a certain number of malicious nodes (to guarantee security). Regarding privacy, a new private data access problem for coded sharding in blockchain systems has been introduced. We plan to use our background in privacy-preserving technology to provide better solutions for coded sharding in blockchain systems when it comes to throughput efficiency, privacy, and security.

Distributed ML: Federated Learning (FL) and Decentralized Learning (DL)

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, distributed ML, e.g., FL and DL, were 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 costs among entities. On the other hand, deep neural networks have shown to be remarkably effective for many ML tasks. An alternative fascinating topic in ML is to find good solutions for enhancing privacy and security through deep learning algorithms. In particular, we are interested in implementing practical privacy-preserving and secure 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.

2 Quantum Information and Computation

Reliable and Efficient Noisy Intermediate Scale Quantum (NISQ) Systems

Quantum computers with 300–1000 qubits, which are referred to as the NISQ devices, may allow us to perform classical tasks beyond the capabilities of today's modern non-quantum digital devices. However, noise in quantum gates is known to be unavoidable, and this intrinsic fact limits the size of quantum circuits that can be reliably executed. While most theoretical results in quantum information and computation are developed under the assumption that quantum resources are unlimited, we are interested in what can be achieved with limited quantum resources in the current realistic NISQ era. We are interested in studying the finite-length analysis and the practical performance of quantum information and computation and aim to develop efficient and reliable quantum information and computation systems using the principles of classical/quantum information theory (IT) and coding theory (CT). This will guarantee the robustness of NISQ systems to perform information processing tasks. We are currently investigating reliable and efficient quantum coding solutions for communicating either classical or quantum information over quantum communication systems with finite quantum resources.

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 IT and CT will be required. Recently, another focus area of our group has been 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.

3 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 a rather general form. The particular details will be discussed and agreed upon 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 are 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 an implementation nature, decent programming skills are essential.

4 About Us

The Master thesis will be supervised by Dr. Eirik Rosnes (leader of the Information Theory Department) and Dr. Hsuan-Yin Lin (senior research scientist) with different degrees of involvement based on the particular project. We have solid experience in IT and CT, as well as many related areas. Currently, our research focus includes various topics in PIR, PC, privacy- and security-enhancing technologies, NISQ systems, and QDISs in general. We mostly consider the research problems from the point of view of IT.

Contact Details

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