Research Industry Talks Schedule

Zoom link (Tuesday):

https://zoom.us/j/7559431931?pwd=TTRHbVc4QU1IODNJNFZmOUw4eElmUT09&omn=98227256842
Meeting ID: 755 943 1931
Passcode: Lb2sMv

==

Tuesday 9.30-9.55: Cagri Polat

Abstract:

In this session, the aim is to briefly outline the General Data Protection Regulation's (GDPR) short history, its purpose, and the essential sub-projects required for implementing data protection compliance projects in both local and international companies. This will include an overview of the market's expectations within a general framework.

Tuesday 9.55-10.20: Abdullah Talayhan

Abstract:

In this talk, we will explore how digital signatures are utilized in decentralized public ledgers like Bitcoin, which heavily relies on the ECDSA signature scheme which is broken by quantum adversaries. To mitigate this attack, bitcoins can be paid to the hash of a public key (P2PKH). Some problems remain with this approach: the owners are vulnerable against rushing adversaries between the time the signature is made public and the time it is committed to the blockchain. Additionally, there is no equivalent mechanism for threshold signatures. We will explore possible directions for solving both of these issues.

==

Zoom link (Thursday):
https://zoom.us/j/7559431931?pwd=TTRHbVc4QU1IODNJNFZmOUw4eElmUT09&omn=98448897270

Meeting ID: 755 943 1931
Passcode: Lb2sMv

==

Thursday 13.30-13.45: Aqsa Shabbir

Abstract:

In industries like finance and healthcare, dealing with large datasets is essential but challenging due to computational constraints and strict privacy regulations. To address these issues, we integrate split learning with homomorphic encryption (HE). Split learning decentralizes deep neural network training by distributing model layers between client devices and a server, thereby alleviating the burden of centralized processing. However, concerns persist regarding potential privacy breaches through the plaintext communication of intermediate outputs and gradients. This is where homomorphic encryption steps in, offering a cryptographic protocol that enables computations on encrypted data without compromising confidentiality. By strategically encrypting only the server side of the model, we aim to mitigate these concerns while leveraging the computational efficiency of split learning. Through optimized task allocation and computation strategies, this integration not only ensures end-to-end security but also minimizes computational overhead. This provides a way for robust and privacy-preserving machine-learning solutions in finance, healthcare, and other data-sensitive domains.

Thursday 13.45-14.00: Kousar Kousar

Abstract:

The sharing of genome data while preserving privacy poses a significant challenge to the advancement of scientific research in the era of big data genomics. To address this, a community-driven protocol known as the genomic data-sharing beacon protocol has gained widespread adoption. This protocol seeks to facilitate secure, easily implemented, and standardized data sharing by permitting only yes/no queries regarding the presence of specific alleles in the dataset. However, despite its intentions, the beacon protocol has been found vulnerable to various privacy breaches, including membership inference and genome reconstruction attacks.

In this paper, we intend to show that privacy threats against genomic data-sharing beacons are not limited to only these attacks, and we uncover a previously unidentified vulnerability: beacon reconstruction attack. We propose two distinct approaches, Optimization and Decoder techniques, for reconstructing entire genomic data-sharing beacons. These methods are predicated on the attacker possessing knowledge of the allele frequencies within the beacon. We aim to illustrate how attackers can exploit SNP correlations to mount efficient attacks, highlighting the broader spectrum of privacy threats facing genomic data-sharing protocols.

The outcome of this attack underscores the importance of continuously evaluating and enhancing the security and privacy measures of genomic data-sharing protocols to mitigate potential risks and protect individuals' sensitive information.

Thursday 14.00-14.20: Melih Cosgun

Abstract:

Data normalization is a very popular and key aspect of training a machine learning model pipeline. Federated Learning is a new and promising area for training ML models without sharing the data, thus preserving privacy. However, applying these data normalization techniques to distributed data is challenging because it requires information about the whole data. Our work mainly consists of two parts. Part 1 is analyzing the effects of different normalization techniques in different Federated Learning scenarios. Part 2 is to apply these normalization techniques in Federated Learning while preserving privacy. To overcome this problem, we use Multiparty Homomorphic Encryption, which allows computations on encrypted data.

Thursday 14.30-14.55: Sylvain Chatel

Abstract:
Homomorphic Encryption (HE) enables computations to be executed directly on encrypted data, without decryption. As such, it is an auspicious solution for protecting the confidentiality of sensitive data without impeding its usability. However, HE does not provide any guarantees that the cryptographic material used has been honestly generated and that the computation was executed correctly on the encrypted data. As such, even though many practical systems rely on HE to achieve strong privacy guarantees, they consider only an honest-but-curious threat model in their constructions.
Therefore, in this talk, we propose solutions to protect different parts of the HE pipeline against malicious adversaries without compromising on the HE scheme.

Thursday 14.55-15.20: Gamze Gursoy

Abstract:
There are major efforts underway to make genome sequencing a routine part of clinical practice. A critical barrier to these is achieving practical solutions for data ownership and integrity. Blockchain provides solutions to these challenges in other realms, such as finance. However, its use in genomics is stymied due to the difficulty in storing large-scale data on-chain, slow transaction speeds, and limitations on querying. To overcome these roadblocks, we developed a private blockchain network to store genomic variants and reference-aligned reads on-chain. It uses nested database indexing with an accompanying tool suite to rapidly access and analyze the data.