Problem Statement Evaluation - POIS

TEAM: RISK RANGERS

Literature Review and Key Ideas

Our literature review unearthed significant advancements in fields crucial to enhancing data privacy:

- "Blockchain technology introduces methods like ring signatures and zero-knowledge proofs to protect transaction privacy."
- "In the realm of cryptography, innovations such as anonymous authenticated encryption (anAE) preserve sender anonymity."
- "Differential privacy and its application across sectors show promise in protecting client data while enabling statistical analysis."
- "Technologies safeguarding anonymity in digital communications, like enhancements to Tor, are vital in the ongoing battle against privacy breaches."

Privacy-Preserving Transactions With Verifiable Local Differential Privacy

The research paper introduces a privacy-focused transaction method for blockchain systems, enabling token exchange with confidential attributes to safeguard user privacy. It incorporates verifiable local differential privacy techniques to ensure data integrity and privacy preservation even in high network usage scenarios. Operating within a permissioned setting, it prevents manipulation of user attributes, addressing concerns related to privacy and data integrity. The scheme allows users to exchange tokens anonymously, while data analysts can collect aggregated information without identifying specific individuals, focusing solely on statistical insights. Leveraging protocols for randomness and advanced privacy-preserving transfer algorithms, along with zero-knowledge Succinct Non-interactive ARgument of Knowledge (zk-SNARKs), ensures the accuracy of differentially private query outputs. Overall, the paper presents a comprehensive framework for privacy-preserving transactions within blockchain systems, addressing challenges concerning privacy, data integrity, and user trust through cryptographic techniques and carefully designed components.

Correlated-Output Differential Privacy and Applications to Dark Pools

My research paper introduces the concept of round differential privacy within the trusted curator model, where a trusted curator evaluates functions on privately submitted inputs and returns outputs privately to clients. To enhance client privacy and prevent leakage, it proposes correlated-output differential privacy. Moreover, the research paper suggests round-differentially-private market mechanisms for both traditional finance and decentralized finance, aiming to counter front-running in market transactions.

Proposed market mechanisms such as volume matching of orders and double auctions are designed to adhere to round differential privacy principles. Additionally, the paper explores the use of secure multi-party computation (MPC) to distribute the trusted curator and implement these algorithms, with a focus on achieving practical feasibility and high throughput for real-world applications.

Key contributions include the introduction of a definitional framework for differential privacy in the trusted curator model, as well as the proposal of round differentially private market mechanisms. It outlines specific algorithms and phases, such as the rDP-Volume-match algorithm, to address privacy challenges effectively. It emphasizes the importance of efficient implementations using custom MPC protocols to facilitate the practical deployment of round-differentially-private mechanisms. Additionally, the paper provides references to related works and additional resources for deeper exploration and research.

Anonymous Cryptocurrencies

Anonymous cryptocurrencies have gained significant attention due to their ability to address the privacy concerns in traditional blockchains like Bitcoin, where addresses are public and susceptible to tracking analysis. Two notable cryptocurrencies in this field are Monero and Zcash, each offering unique solutions to enhance privacy and security.

Monero employs ring signatures and stealth addresses to obfuscate transaction details, making it difficult to trace the sender and receiver. Additionally, Monero’s dynamic block size algorithm ensures scalability while maintaining privacy. However, one drawback is the continuously growing set of potential unspent coins, leading to challenges in storing a concise representation of the blockchain.

Zcash utilizes zero-knowledge proofs, specifically zk-SNARKs, to provide selective transparency. Users can choose between shielded and transparent transactions, balancing privacy and auditability. Despite offering strong privacy features, Zcash faces criticisms regarding its trusted setup and reliance on zk-SNARK technology.

While there exist cryptocurrencies that address the privacy concerns inherent in blockchain, they are not without their pitfalls. There is a need for a cryptocurrency that achieves provably secure notions of anonymity, while maintaining a concise representation of the blockchain.

Untagging Tor: A Formal Treatment of Onion Encryption

The literature review explores the foundational concepts of anonymity in digital communication, tracing its origins from mix-nets to the development of onion routing protocols like Tor. It highlights Tor's significance in providing secure, anonymous internet access while discussing its vulnerability to tagging attacks.

The introduction sets the stage by elucidating Tor's architecture and the encryption mechanisms it employs to safeguard user identities. Despite its robust cryptographic techniques, Tor faces threats from tagging attacks, prompting the Tor community to propose enhancements to its encryption scheme.

The paper delves into Proposal 261, which aims to mitigate tagging attacks by enhancing Tor's onion encryption. It adopts a formal treatment of circuit-based onion encryption, offering insights into the complexities of balancing routing functionality with security considerations. The analysis validates Tor261's design but also identifies potential vulnerabilities, particularly in its support for RELAY EARLY cells.

Overall, the literature review underscores the ongoing challenges in securing anonymous communication networks like Tor. It emphasizes the need for continuous research and development to address emerging threats and maintain the balance between security and usability in privacy-enhancing technologies.

Anonymous AE

Through a thorough analysis of current cryptographic methods, it becomes apparent that traditional authenticated encryption (AE) techniques possess inherent privacy vulnerabilities. In the standard AE framework, decrypting a message necessitates the knowledge of several components: a nonce (a number utilized only once), potentially a session identifier (SID), and any associated data (AD) used during the encryption process.

These elements are typically conveyed alongside the encrypted message, often without any form of protection. This conventional approach poses a risk to the sender's anonymity, as interceptors of the communication could potentially glean insights about the sender or the session from these unencrypted pieces of information.

In response to these privacy concerns inherent in traditional AE, the literature introduces the concept of anonymous AE (anAE). anAE ensures the privacy of the sender's identity by ensuring that all necessary decryption information, barring the receiver's private keys, is contained within the ciphertext. A key innovation within this sphere is the Nonce Wrap method, a specific technique designed to convert a standard nonce-based AE scheme into an anAE scheme. Nevertheless, while anAE effectively addresses privacy concerns related to the message content, it does not fully mitigate vulnerabilities related to traffic-flow analysis.

By integrating all required information within the encrypted message itself and ensuring anonymity, anAE substantially enhances both the privacy and usability of encrypted communications. This exploration underscores the critical need for ongoing research and innovation in cryptography to address not only the privacy concerns of today but also the sophisticated threats of tomorrow.

ACADEMIC TRENDS

The literature review we conducted for the project proposal helped us to find out more about significant advancements and methodologies in the field of data privacy, particularly in the realms of cryptography, blockchain privacy, differential privacy, and anonymous communication networks. These key ideas lay the groundwork for addressing privacy concerns across various domains, including healthcare, finance, and digital communication:

- **Advancements in Cryptocurrency Privacy:** Studies on Monero and Zcash highlight the pursuit of enhanced privacy in blockchain transactions. The use of ring signatures, stealth addresses, and zero-knowledge proofs (zk-SNARKs) showcases innovative approaches to maintaining transaction privacy while addressing scalability and auditability challenges.

- **Innovations in Anonymous Authenticated Encryption:** The exploration of anonymous AE (anAE) and the NonceWrap method underscores the importance of preserving sender anonymity in encrypted communications. This addresses the need for encryption techniques that protect against traffic-flow analysis and other forms of privacy breaches.

- **Differential Privacy for Enhancing Client Confidentiality:** The concept of round differential privacy within the trusted curator model presents a novel approach to safeguarding client data in both traditional and decentralized financial systems. This includes mechanisms to prevent front-running and leakage, emphasizing the importance of privacy-preserving market mechanisms.

- **Protecting Anonymity in Digital Communications:** Research into securing networks like Tor against tagging attacks illustrates the ongoing challenges in ensuring user anonymity online. Proposals for enhancing encryption schemes indicate the critical need for constant evolution in technologies that protect user identities and data.

- **Blockchain Privacy-Preserving Transactions:** The development of methods for statistical data collection within blockchain systems, without compromising individual privacy, showcases an innovative approach to maintaining data integrity. The use of verifiable local differential privacy and zk-SNARKs points to sophisticated cryptographic solutions for privacy preservation.

RESEARCH INTEREST

- The problem that captivates our interest revolves around the critical issue of data privacy for organizations possessing valuable datasets containing personally identifiable information (PII) and sensitive personally identifiable information (SPII). Specifically, we are drawn to the challenge of anonymizing datasets in a manner that adheres to privacy regulations like HIPAA while still enabling the utilization of advanced data processing and machine learning tools offered by external entities.

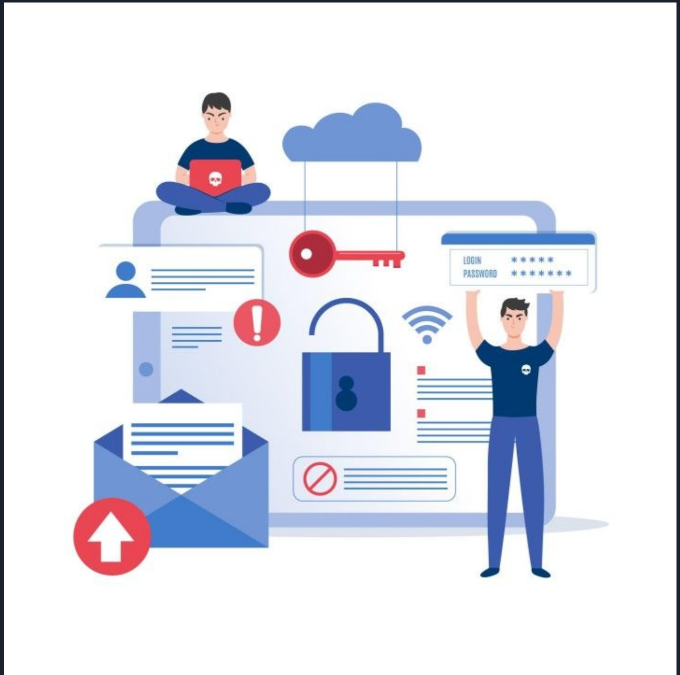
- The existing dilemma for sectors such as healthcare, finance, and education, among others, is their inability to leverage cutting-edge technologies for fear of privacy breaches. Our project, therefore, aims to address this gap by proposing "Ashe," a data anonymization tool. Ashe is designed to utilize state-of-the-art differential privacy algorithms to anonymize datasets before their public release for statistical analysis. This tool seeks to ensure that individual identities are protected through data modification techniques that maintain statistical significance, thereby enabling organizations to safely share their data for broader analytical purposes without compromising individual privacy.

- This problem is not only intellectually stimulating but also highly relevant in today's data-driven world, where the balance between data utility and privacy is of paramount importance.

Challenge of Anonymizing Data

- Organizations possess valuable datasets that, if analyzed, could unlock insights into human behavior, market trends, and more. Yet, the presence of PII and SPII creates a privacy issues.

- Regulations like HIPAA pose a significant challenge, restricting the use of advanced data processing tools due to the risk of privacy breaches. This results in a substantial underutilization of potential insights that could be gained from such data.



PROPOSED PROBLEM STATEMENT

Introducing Ashe - Our Proposed statement

- To address these challenges, we propose 'Ashe' - a cutting-edge data anonymization tool which employs state-of-the-art differential privacy algorithms to transform datasets.

- By tweaking data attributes while preserving statistical significance, Ashe enables the safe sharing of anonymized data. This ensures privacy compliance and opens the door to advanced analytical opportunities.

Threat Model - Perfect Anonymization

would imply that an adversary would not gain any additional information from the anonymized dataset. However, it is impossible to construct a threat model around perfect anonymization for differential privacy. This is implicit due to the constraint that the anonymized values must be somewhat close to the true values if we want to maintain statistical significance.

Network Assumptions - Data Transfer Security

We assume that the data transferred over the network, specifically when the organization uploads the true dataset to the Ashe application, is secured using encryption protocols such as HTTPS. In other words, the adversary has no access whatsoever to the true dataset.

Adversary Goal :

If the adversary, with full access to the anonymized dataset, can guess the true identity of any of the records, we declare that the adversary succeeds.

Client Server Architecture :

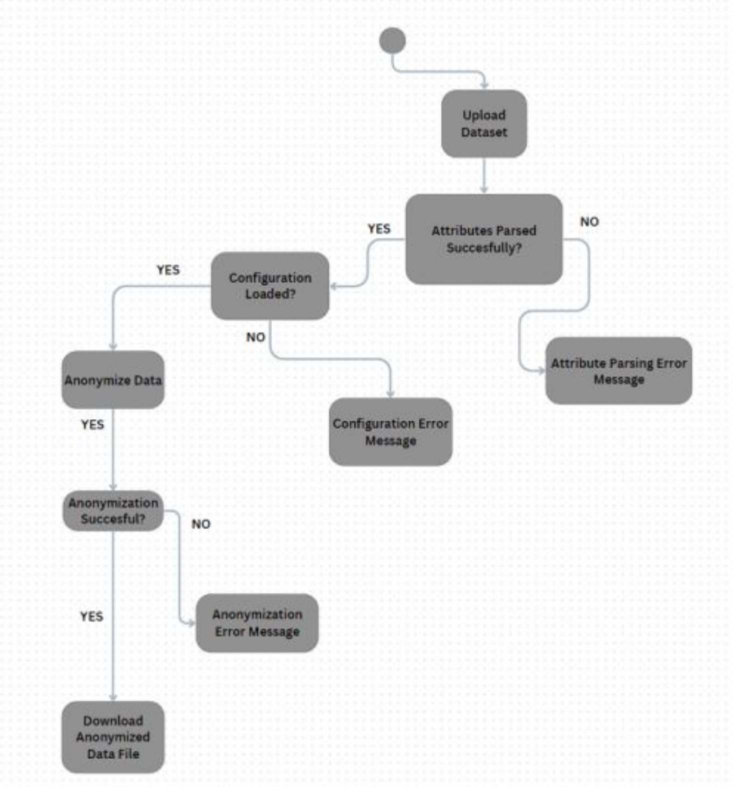
Ashe will be designed with a client-server architecture, where the anonymization tool runs on the client side, and interacts with a server which provides differential privacy algorithms, ability to share data sets etc.

Technical Framework/Architecture

- Ashe is designed with a robust client-server architecture. At its core, it utilizes differential privacy algorithms to anonymize data effectively.

- We prioritize security in data transfer, ensuring encryption protocols like HTTPS are in place to protect the dataset during upload to Ashe.

- Our threat model acknowledges that perfect anonymization is unattainable. However, Ashe aims to make it virtually impossible for adversaries to identify individuals within anonymized datasets.


Upload Dataset
Attributes Parsed Succesfully?
YES
NO
Configuration Loaded?
YES
NO
Attribute Parsing Error Message
Anonymize Data
Configuration Error Message
YES
Anonymization Succesful?
NO
Anonymization Error Message
YES
Download Anonymized Data File

ESSENTIAL READINGS

ESSENTIAL READINGS

Privacy-Preserving Transactions With Verifiable Local Differential Privacy

The paper presents a method for private transactions on blockchain, ensuring user privacy while allowing data analysts to gather aggregated insights. By using verifiable local differential privacy, it maintains data integrity even during high network usage. Operating within a controlled environment, it prevents user information manipulation, addressing privacy and data reliability concerns. This approach offers a secure solution for transactions, balancing privacy and statistical analysis effectively in blockchain systems.

https://eprint.iacr.org/2023/126.pdf

Correlated-Output Differential Privacy and Applications to Dark Pools

The paper introduces round differential privacy in the trusted curator model, enhancing client privacy. It proposes market mechanisms for traditional and decentralized finance to combat front-running. Algorithms like rDP-Volume-match are outlined for efficient implementation using secure multi-party computation. Key contributions include a definitional framework for privacy and practical deployment strategies.

https://eprint.iacr.org/2023/943.pdf

Anonymous AE

Monero and Zcash are prominent anonymous cryptocurrencies addressing privacy concerns in traditional blockchains like Bitcoin. Monero ensures anonymity through ring signatures and stealth addresses, while Zcash offers selective transparency using zk-SNARKs. However, both face challenges such as storage issues and reliance on trusted setups. A cryptocurrency with provably secure anonymity and efficient blockchain representation remains a key goal.

https://eprint.iacr.org/2019/1033.pdf

PROJECT PLAN

(Step-1)

- Setup basic connection between the React frontend and Flask backend
- Developing the frontend with the following pages
- Source - upload original dataset
- Configure - set epsilon values
- Result - display anonymized dataset

(Step-2)

- Integrate PyDP library into the Flask backend
- Ensure that the result of the differential privacy algorithms is accessible to the frontend.

(Step-3)

- Testing - handling large datasets, number of attributes, response time etc.
- Security - scan the codebase for potential security vulnerabilities
- Deployment - dockerize the application.

Thank You

Santhoshini thota (2021101097)

Pradyumn Shukla (2022201001)

Abhay Patil (2020101022)

Pranjal Chapriyal (2020101108) (gay)

Ishank (2023701012)