RESEARCH STATEMENT

Yanxue Jia, Purdue University

My research employs applied cryptography and distributed trust to advance the security and privacy of real-world collaboration and communication systems. My current work focuses on the following two directions.

First, I address fundamental privacy issues prevalent across data collaboration and network communication scenarios. These scenarios often involve participants who do not fully trust each other, as the risk of privacy leakage. A cryptographic technique, called secure multi-party computation (MPC), allows multiple parties to perform joint computations without leaking their own data. My research advances and implements secure and efficient MPC techniques to effectively solve these real-world privacy issues.

Second, I focus on providing specific privacy-preserving solutions for emerging blockchain applications. While blockchains have gained significant attention for their decentralization, transparency, and immutability, these characteristics introduce unique privacy challenges that existing cryptographic primitives and security models cannot address. To address these challenges, I develop novel cryptographic primitives and establish new security models that enhance privacy within blockchain ecosystems.

In the near future, besides further delving into the current research, I also plan to advance cryptographic techniques through *hardware-accelerated* designs. My long-term goal is to develop privacy-enhancing solutions for compute-intensive applications.

1 Privacy Protection via Multi-Party Computation

My current research focuses on using two-party secure computation to address the fundamental privacy issues in data collaboration and network communication scenarios. Specifically, my work [JSZ+22, JSZG24] proposed a unified framework for private set operations and especially improved the efficiency and security of private set union (PSU), and my work [JMK24, JDZK] designed server-aided metadata-hiding communications, comprehensively meeting the requirements in practice.

(1) Private Set Operations (PSO). A large number of applications require set operations among different parties. Directly revealing the individual sets poses privacy risks, especially for privacy-sensitive parties, such as banks and hospitals. Private set operations enable participants to obtain only the desired set operation result without revealing any other information about their sets. Specifically, private set intersection (PSI) is needed in scenarios where only the intersection is required and other elements must remain protected, e.g., contact discovery, password breach alerting, and conversion rate of advertisements. Conversely, Google maintains a project called "Private Join and Compute"[1], which hides the intersection but obtains the results of further data processing (e.g., sum, average, and cardinality) on the intersection. Likewise, private set union (PSU) also protects intersection but allows the participants to obtain the union, which can be used to compute joint IP blocklists to enhance the accuracy of identifying malicious sources and compute the union of training samples for federated learning.

Unified Framework for PSO. Previous work focused on a single type of private set operations, which is not sufficient for scenarios requiring multiple types of set operations and increases the deployment workload. My work [JSZ+22] proposed a new framework, under which I obtained a more efficient PSU protocol. Moreover, the framework can also be used to unify the above private set operations. The key difference among these private set operations lies in whether the intersection is hidden. The framework enables flexible control over whether the intersection remains hidden, by using a novel building block called "Generalized Reverse Private Membership Test (GRPMT)". GRPMT separates an element and its corresponding membership test result between the two parties, and then specific utilization of these membership test results determines the type of set operation. Furthermore, I utilized lightweight symmetric-key techniques to

[1] Private Join and Compute, Google, https://github.com/google/private-join-and-compute.git.

achieve GRPMT, avoiding the heavy computation required in previous hidden-intersection designs. *This work has transitioned PSO from theoretical exploration to practical deployment, catalyzing a wave of subsequent research and innovation in the field, as evidenced by 37 citations since August 2022.*

PSU with Stronger Security. My recent work [JSZG24] uncovered a previously overlooked leakage issue in existing efficient PSU protocols, which I defined as during-execution leakage. I demonstrated that this leakage can empower adversaries to obtain additional information in real-world scenarios. My comprehensive analysis of existing PSU designs revealed that only those based on computationally expensive homomorphic encryption are immune to this issue, but these solutions are impractical for real-world use. To address this gap, I developed a new specific PSU framework that avoids the leakage issue and instantiated this framework using lightweight cryptographic techniques, making it both secure and highly efficient. In addition, I revealed that the current PSU security definition cannot be used to identify the leakage. To prevent future PSU designs from having such leakage, I provided an enhanced security definition for PSU, such that the designs suffering from the leakage cannot be proved secure using my security definition. My work provides a strong theoretical foundation for future research on PSU.

(2) Server-aided Metadata-Hiding Communication. It is well understood that revealing the content of communications poses a privacy risk. However, the dangers of leaking metadata—that is, information concerning who communicated with whom, when, and the extent of their interactions—are often overlooked. In reality, metadata leakage alone can also endanger the communicating parties. For instance, consider a whistleblower: if their act of whistleblowing is exposed, even without disclosing the specific content of their report, they could still face serious threats to their life, as confirmed by numerous reports. Metadata-hiding communication plays a crucial role in providing strong security and privacy guarantees for metadata. Server-aided metadata-hiding communication has gained increasing attention due to its ability to reduce costs for both senders and receivers. In essence, this technique allows servers, senders, and receivers to securely complete a communication task—enabling interactions between senders and receivers while ensuring that metadata remains concealed.

An ideal server-aided metadata-hiding communication system is expected to meet the three functional requirements: (1) the number of messages that a receiver is allowed to receive is unlimited; (2) the sender can send a message to a receiver even if the receiver is not currently online, and the receiver is allowed to retrieve the messages when coming back; (3) the retrieved messages can be securely deleted from the server(s). None of the previous designs achieved the three requirements simultaneously. My work [JMK24] focused on protecting the receiver's privacy and achieved the three requirements simultaneously for the first time. Moreover, compared to previous solutions, our performance showed an improvement of over $1000\times$. Furthermore, my work can be used not only for traditional communication scenarios, but also for blockchain scenarios where senders publicly post their messages on a bulletin board (i.e. a blockchain) and recipients retrieve their messages from the bulletin board. This process involving blockchain presents more opportunities for various intermediaries to potentially link senders to recipients by tracking these accesses. My current work [JDZK] further protected the sender's privacy, achieving an end-to-end metadata-hiding messaging system that meets the above three functional requirements.

[Future Research]

In the near future, I will extend my work on private set operations and server-aided metadata-hiding communication. Specifically, I have been working on fuzzy Private Set Intersection (PSI), which is a generalization of PSI. Different from PSI testing if two elements are exactly equal, fuzzy PSI evaluates if their similarity exceeds a predefined threshold, which is applicable in fingerprint and iris scans, ride-sharing platforms, and illegal content detection. In addition, in real-world scenarios, data sets are often dynamic; therefore, updatable PSO is also an area I plan to research in the future. While metadata-hiding communication provides strong privacy for users, it also creates opportunities for malicious parties, as it becomes difficult to track those who post illegal messages. Therefore, striking a balance between preserving anonymity and track-

ing malicious behavior is a challenge I plan to tackle in the future. Furthermore, I will also explore other important specific tasks, such as oblivious sorting, secure aggregation, and graph computation.

2 Privacy-Preserving and Regulation-Compliant Blockchains

Blockchains provide strong support for decentralization, primarily because on-chain data is immutable and can be publicly accessed. However, these characteristics also expose blockchain users to more severe privacy threats and make it more challenging to deploy regulatory-compliant systems. *My research addresses these privacy vulnerabilities in blockchain systems [JSZ+20, JSZG22] and redesigns the blockchain architecture to ensure that it remains both legally compliant and decentralized [JSZ+21, LJGK].*

(1) Privacy-Preserving Payment System. One of the main applications of blockchain is distributed payment systems (namely, cryptocurrencies), where users' transaction data are public, and thus can be traced. Ring signature and coin shuffling are two typical techniques used to achieve privacy-preserving payments, and the core idea is to use an anonymity set to hide the real participants of a transaction. For each transaction, the approach based on ring signature generates a ring signature, which will be stored on the blockchain. The signature size of previous works is linear in the anonymity set size. My work [JSZ+20] reduced the signature size to logarithm in the anonymity set size, which means that when achieving the same anonymity level, the performance of my solution is much better. On the other hand, the previous works based on the coin shuffling technique always needed interactions between participants, which forces users to remain online. My work [JSZG22] designed a non-interactive payment system only relying on lightweight cryptographic techniques, enabling users to transact asynchronously without needing to keep online, thereby enhancing both convenience and efficiency. Additionally, I for the first time established a security definition in the real/ideal simulation paradigm for the non-interactive payment system.

(2) Regulatory-Compliant Storage System. Research has revealed the presence of substantial illegal data on blockchain systems, including copyright-infringing files, privacy-violating images and information, and links to illegal websites. Additionally, the General Data Protection Regulation (GDPR) of the European Union gave individuals the right to delete their personal data (i.e., "Right to be Forgotten"). However, immutability is crucial to the security of blockchain applications, and altering this characteristic without compromising security is not a trivial task. Previous approaches either introduced a powerful regulator with full control over on-chain data or allowed users to modify data at will, opening the door to potential malicious actions. My work [JSZ+21] obtained a balance between regulation and autonomy. My approach allows users to manage their own data, while giving regulators the authority to remove illegal content. Crucially, any modifications initiated by the regulator are only validated upon approval by the majority of users, ensuring a secure and democratic process.

(3) Securely Feeding TLS-Protected Data to Blockchains. Smart contracts executing on blockchains commonly require accessing data (e.g., age and bank balance) from external websites beyond the blockchain. However, smart contracts lack network access to these external websites. On the other hand, Transport Layer Security (TLS) can ensure that a client securely retrieves data from a website, but it does not allow the client to prove to a third party that the data originates from the website. Therefore, a service, called Oracle, is needed to securely feed TLS-protected data to blockchains while protecting the users' private information (e.g., passwords). The previous works always involved secure computations or zero-knowledge proofs, which incur significant computation and communication costs.

My recent work [LJSK24] dived into the security of the Authenticated Encryption with Associated Data (AEAD) schemes used in TLS 1.2/1.3. Surprisingly, I found that, in two highly prevalent scenarios, the expensive secure computations or zero-knowledge proofs can be completely eliminated and the service only needs to forward messages between the client and website. The two scenarios are: (1) when HTTP is used as the application layer protocol, and (2) when the client cannot modify data stored on the website. At the

same time, I introduced a new security property for AEAD, addressing a previously unrecognized gap in its security analysis.

[Future Research]

In my future research on blockchains, I will prioritize both privacy and decentralization as guiding principles. Privacy-preserving features, while essential, often increase miners' burden, raising entry barriers and potentially undermining decentralization. To address this, our recent work [LJGK] introduced an updatable and aggregatable vector commitment, enabling stateless cryptocurrencies by allowing miners to maintain only a succinct digest of the current state. Integrating this commitment with privacy-preserving payment solutions will be a focus of my future exploration. I also plan to explore the use of multiple servers to develop Oracle services based on MPC techniques, reducing dependency on a single Oracle node. Furthermore, I plan to expand the privacy-preserving solutions for transactions to other areas, to address privacy issues across different application scenarios, such as digital assets management and healthcare.

3 Emerging Research: Hardware-Accelerated Cryptography

In the future, I plan to revisit the design of cryptographic primitives from a hardware compatibility perspective, to enable hardware acceleration for some advanced cryptographic techniques. A significant amount of work has begun to explore the use of advanced cryptographic tools—such as MPC, Zero-Knowledge Proof (ZKP), and Fully Homomorphic Encryption (FHE)—in complex application scenarios. For example, in the machine learning area, MPC and FHE can be used to achieve privacy protection and ZKP can be used to guarantee correct model computations. However, the performance is still far from practical uses.

As a first step in this direction, I have been working on the integration of AES (Advanced Encryption Standard) primitives into GPU architectures. In the field of MPC, key components—such as Garbled Circuits (GC), Distributed Oblivious Random Access Machine (DORAM), and Function Secret Sharing (FSS)—rely heavily on numerous AES operations. Moreover, these components are sufficient to support various types of computations: (1) GC enables the secure evaluation of arbitrary functions; (2) a combination of GC and DORAM offers more efficient MPC solutions for RAM programs, including database queries, graph algorithms, etc; (3) a novel FSS-based MPC approach has been developed to handle non-arithmetic operations (e.g., truncation, integer comparison, and ReLU) more efficiently.

Our experiments demonstrate that integrating our GPU-accelerated AES into these components significantly reduces computation costs. However, other parts of the process (e.g., memory I/O), which have not been optimized, now present new performance bottlenecks. Moving forward, I plan to address these new bottlenecks by refining cryptographic designs (e.g., reducing data size in memory I/O operations). On the other hand, I also plan to maximize the use of AES in designing novel MPC solutions (e.g., securely replacing hash functions with AES), fully utilizing the benefits provided by our GPU-accelerated AES.

Integrating AES into GPU architectures and applying it to MPC is only the beginning. Also, I will explore hardware-accelerated ZKP and FHE. A major focus will be accelerating the number-theoretic transform (NTT), a bottleneck in both ZKP schemes (e.g., Groth and Plonk) and FHE schemes. In addition, the recent ZKP schemes use other modules—such as sum-check protocol, Merkle tree, and linear-time encoder—rather than NTT. Hardware acceleration for these modules also requires deeper research. For FHE, two necessary ciphertext maintenance operations (i.e., key-switching and bootstrapping) are extremely time-consuming, necessitating designing end-to-end hardware acceleration solutions for both operations.

References

[JDZK] **Yanxue Jia**, Debajyoti Das, Wenhao Zhang, and Aniket Kate. Kerblam — Anonymous Messaging System Protecting Both Senders and Recipients. In Submission. https://yanxue820.github.io/files/Kerblam.pdf.

[JMK24] **Yanxue Jia**, Varun Madathil, and Aniket Kate. HomeRun: High-efficiency Oblivious Message Retrieval, Unrestricted. In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2024.

[JSZ+20] **Yanxue Jia**, Shi-Feng Sun, Yuncong Zhang, Qingzhao Zhang, Ning Ding, Zhiqiang Liu, Joseph K Liu, and Dawu Gu. PBT: A new privacy-preserving payment protocol for blockchain transactions. *IEEE Transactions on Dependable and Secure Computing (TDSC)*, 2020.

[JSZ+21] **Yanxue Jia**, Shi-Feng Sun, Yi Zhang, Zhiqiang Liu, and Dawu Gu. Redactable blockchain supporting supervision and self-management. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security (AsiaCCS)*, 2021.

[JSZ+22] **Yanxue Jia**, Shi-Feng Sun, Hong-Sheng Zhou, Jiajun Du, and Dawu Gu. Shuffle-based Private Set Union: Faster and More Secure. In *31st USENIX Security Symposium (USENIX Security 22)*, 2022.

[JSZG22] **Yanxue Jia**, Shi-Feng Sun, Hong-Sheng Zhou, and Dawu Gu. A Universally Composable Non-interactive Aggregate Cash System. In *International Conference on the Theory and Application of Cryptology and Information Security (Asiacrypt)*, 2022.

[JSZG24] **Yanxue Jia**, Shi-Feng Sun, Hong-Sheng Zhou, and Dawu Gu. Scalable Private Set Union, with Stronger Security. In *33st USENIX Security Symposium (USENIX Security 24)*, 2024.

[LJGK] Zhongtang Luo, **Yanxue Jia**, Alejandra Victoria Ospina Gracia, and Aniket Kate. Cauchyproofs: Batch-Updatable Vector Commitment with Easy Aggregation and Application to Stateless Blockchains. In Submission. https://yanxue820.github.io/files/Cauchyproofs.pdf.

[LJSK24] Zhongtang Luo, **Yanxue Jia**, Yaobin Shen, and Aniket Kate. Proxying is Enough: Security of Proxying in TLS Oracles and AEAD Context Unforgeability. *The Science of Blockchain Conference (SBC)*, 2024. https://eprint.iacr.org/2024/733.