Research Statement
Sujaya Maiyya

Individuals and enterprises produce over 2.5 exabytes ($10^{18}$ bytes) of data everyday. Much of this data - including sensitive and private information - is stored with and managed by third parties, such as Amazon Web Services or Google Cloud. These companies can lose millions to billions of dollars in sales if their data access latencies increase by only a few hundred milliseconds [1]. However, there is a fundamental tradeoff between security and efficiency in data management systems. While these companies own and trust their underlying data infrastructure, and so can prioritize latency, the same is not true for those not directly controlling the storage and management of their data. My PhD research focuses on designing, prototyping and evaluating protocols that strike a balance between efficiency and security in both trusted and untrusted systems.

In my recent work, with efficiency as the focus, we developed Samya [2] (ICDE 2021) as a database system that improves latency for high contention, hotspot data compared to existing databases. Samya is a geo-distributed data management system that stores and manages hotspot aggregate data. Samya dis-aggregates aggregate data and stores partitions of the dis-aggregated data on different servers. This design choice allows servers in Samya to independently and concurrently serve client requests. State-of-the-art geo-distributed databases such as Google’s Spanner take a centralized approach where a server acting as a leader processes client requests sequentially. Compared to the centralized solution, Samya’s parallelism reduces the 99th percentile latency by 76% and allows Samya to commit 16x to 18x more transactions.

While Samya successfully iterated on traditional cloud infrastructure I became intrigued by the question, “What if I host my data on completely untrusted infrastructure?”. To answer this question, we recently proposed TFCommit [3] (ICDCS 2020), a trust-free commitment protocol that executes transactions across multiple untrusted servers. To our knowledge, TFCommit is the first atomic commitment protocol to tolerate malicious failures without using replication. TFCommit combines a transaction commitment protocol with collective signatures and produces a tamper-resistant log; this log can then be audited to detect faulty behavior. TFCommit is relevant in many scenarios, such as supply-chain management where the participating entities span different administrative domains that do not trust each other. Compared to executing a transaction in trusted infrastructures, the overhead of executing TFCommit is 1.8x in latency and 2.1 in throughput - an acceptable overhead given the additional security guarantees of TFCommit.

Future research interests:
"FTC imposes $5 Billion penalty on Facebook for violating consumers' privacy"[4]. The increasing number of financial penalties indicate a lack of strict data privacy guarantees from current data management systems. My ongoing projects aim to build a foundation for the next generation of data management systems that preserve privacy while being both fault-tolerant and efficient. Many open problems in databases can be, and in fact must be, expanded to tolerate
malicious behavior to preserve data privacy and security. The goal of my research is to build efficient and practical database systems that meet the increasing public expectations and legal requirements for data privacy that are currently lacking in existing systems.

References: