Although first mentioned in a paper in 1991, blockchain technology has grown in acceptance and importance in recent years. The decentralized ledger of transactions behind cryptocurrencies such as Bitcoin, blockchains also find use in supply chains, advertising, and medical information systems. Blockchains are very secure; each block in the chain contains a hash of the previous block, which makes it almost impossible to tamper with one block because the hash in the other block will no longer match. As blocks are added to the chain however, scalability becomes a serious issue due to the massive volumes of transactional data. The project team sees sketch data structures as the answer to this issue. Sketches represent large data sets and use a fraction of the space of the original data. The data can be queried, but these data structures do not always provide accurate results.

Sketch data structures such as Bloom and cuckoo filters, Merkle trees, and invertible Bloom lookup tables (IBLT) find popular use in querying and proving element membership in a given set. Sketches can also be used, to varied success, in determining set synchronization and similarity, counting, and determination of frequently occurring elements.

For blockchain applications, these data structures and algorithms are currently used for diverse functions, such as block summaries. In this case, a Bloom filter or Merkle tree contains a summary of the block, speeding the search for transactions. Simplified payment verification in Bitcoin uses a light client in association with a Bloom filter, although there are privacy issues. Unique address checking, pool synchronization, and fairness enforcement all use sketches as mechanisms to handle address space issues, set comparisons, and block transaction selection. Finally, Bitcoin and Ethereum use Merkle trees to organize transaction and status information.

In terms of future applications, the project team sees benefit by applying sketches for network analysis and anomaly detection. Counters can be used to identify high-traffic accounts, but this would result in many counters. Instead, sketches used for basic counting tasks can be applied in looking for the most active addresses with the highest number of transactions. The project team suggest implementing sketches beyond the Bloom filter to simplify computations. But they should be limited as they are more memory intensive.

The advances in sketches that were highlighted include the use of sketches for multiple tasks instead of a single counting task, accurate Bloom filters that eliminate false positives, and learned Bloom filters that prevent false negatives. The team has already presented their initial results in a conference paper.