Scalable Privacy-Preserving Query Processing Over Ethereum Blockchain

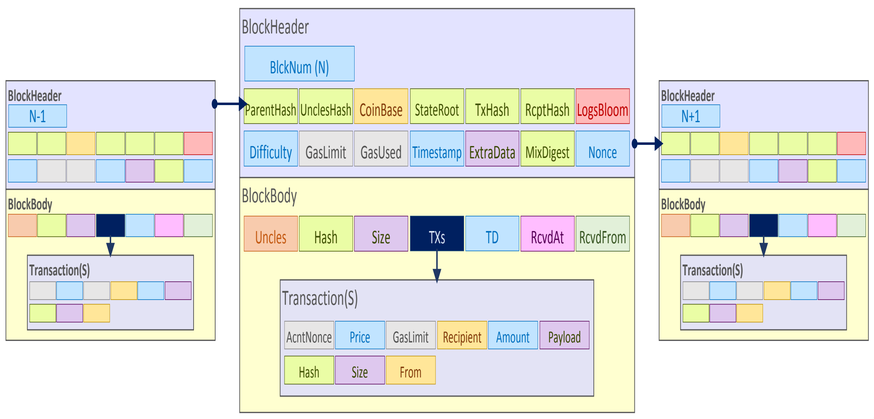
Shlomi Linoy, Hassan Mahdikhani, Suprio Ray, Rongxing Lu, Natalia Stakhanova and Ali Ghorbani

Canadian Institute for Cybersecurity (CIC), University of new Brunswick (UNB)

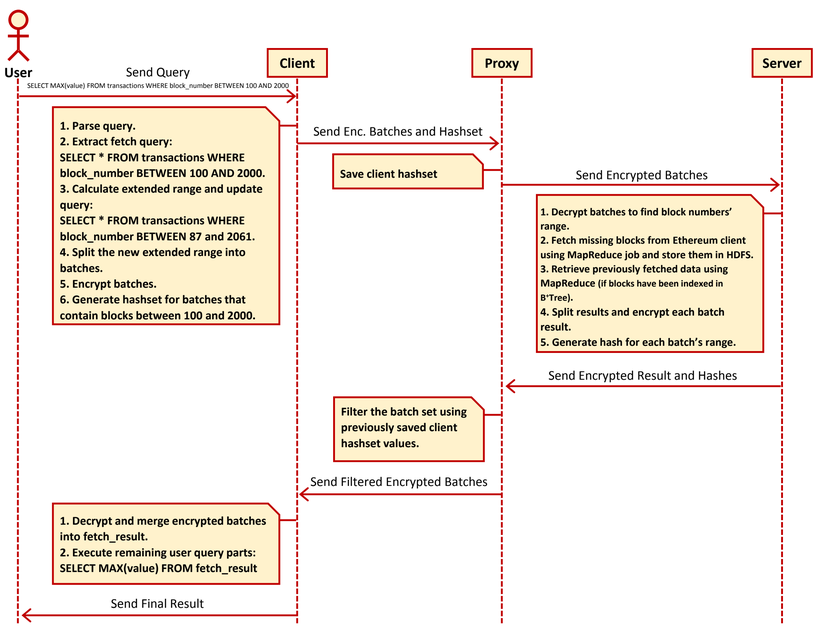
ABSTRACT

Blockchain technologies have recently received considerable attention, partly due to the success of crypto-currency applications such as Bitcoin and Ethereum. As the adoption of blockchain technologies by various sectors increases, there is a need for tools that enable regulation enforcement, for e.g. enabling auditors to monitor, examine and ensure compliance of the data stored by the blockchain systems with regulations in a privacy preserving way. Primarily a transaction repository, blockchain based systems provide very limited query support and extending it could be cumbersome. For instance, the execution of queries over an Ethereum client requires the installation of a client node (e.g. Geth), and its synchronization with the whole Ethereum network can take several days. Furthermore, its current query support (using Web3 API) is limited to basic search capabilities, like searching a block by a single block number or hash code, which can be time consuming and inefficient when a large number of blocks or transactions are required. Finally, there is no support for privacy-preserving query processing within Ethereum clients. To address these issues, we propose a system that can provide auditors with efficient, scalable and richer blockchain query processing over Hadoop and synchronized Ethereum clients. In addition, the system ensures the auditors' privacy by utilizing cryptography techniques over semi-trusted servers to protect the auditors' identities, queries and their results in all involved parties.

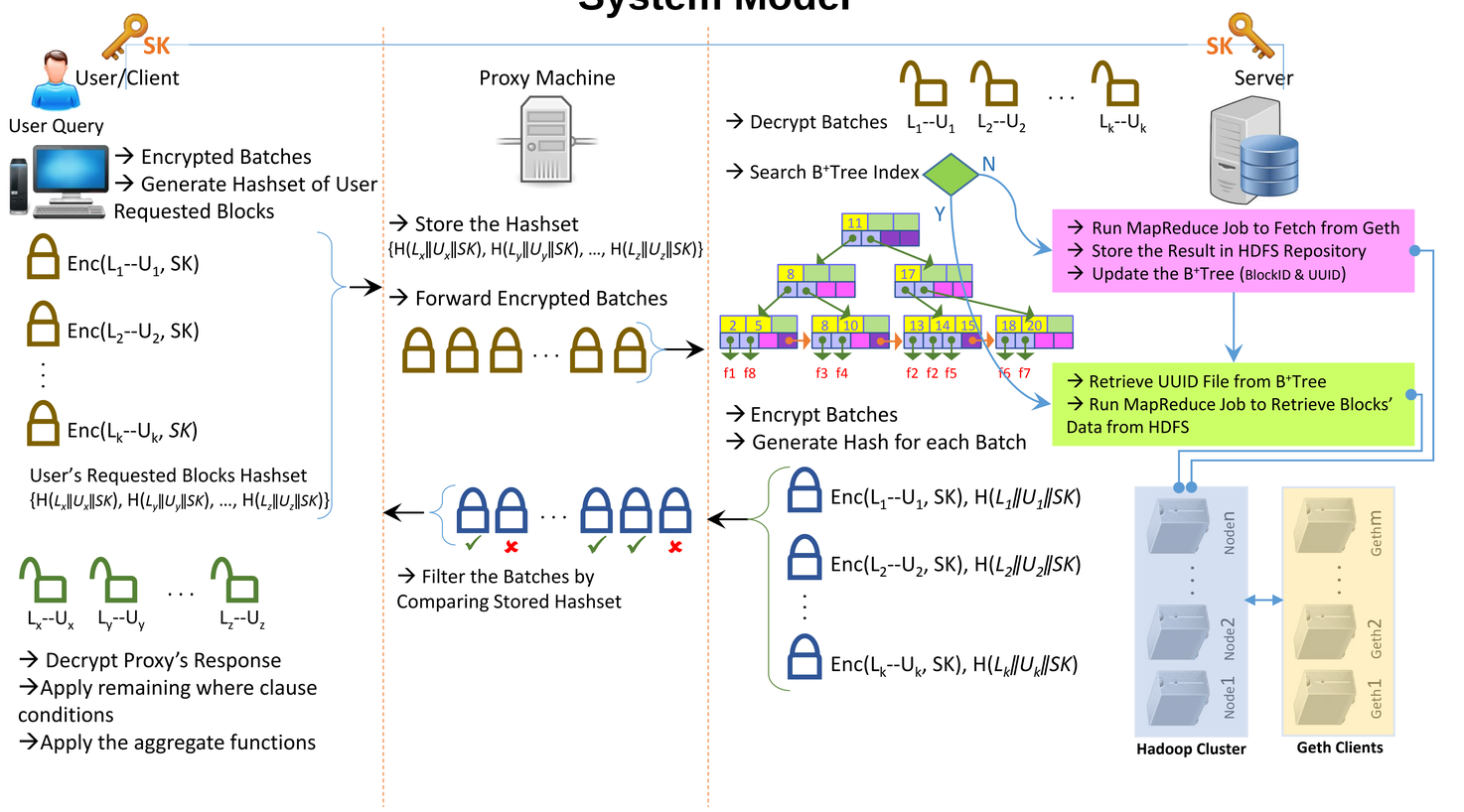
Ethereum's Block/Transaction



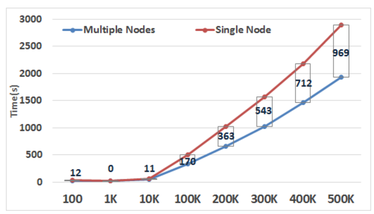
Flow of User Query Execution



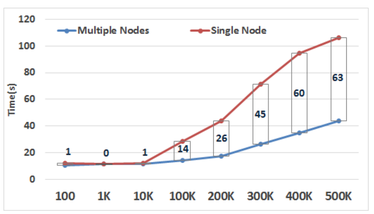
System Model



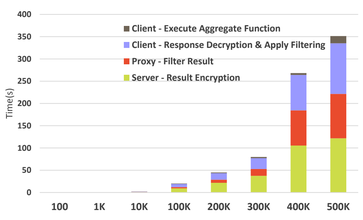
Performance Evaluation



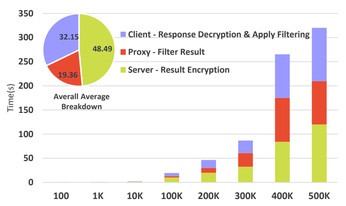
Fetching from Ethereum Geth clients



Fetching from HDFS



SELECT MAX (value)
FROM transactions
WHERE blocknumber
BETWEEN <LB> AND <UB>.



SELECT *
FROM transactions
WHERE blocknumber
BETWEEN <LB> AND <UB> AND
accountaddress = <someaccountaddress>

Conclusion

As an increasing number of sectors integrate blockchaintechnologies, it is important to have an efficient and secureauditing system to help monitor and analyze blockchainrepositories while preserving the auditors' privacy. To thisend, our proposed system uses big data processing tech-niques to support all the above requirements. Thanks to bigdata infrastructures being very affordable and accessible, inaddition to having blockchain repositories transparent bydesign, any private, academic, or commercial entity caneasily implement and use the proposed system. Our systemprovides a secure, robust, and scalable way to process SQLqueries over any blockchain and enables multiple auditorsto execute queries over blockchain data in an efficientand scalable way, while preserving the privacy of auditors'identities and prevent the disclosure of the queries beingused and their results. It supports SQL queries with rangeand aggregate functions. It transforms each SQL queryinto MapReduce tasks to be run using Hadoop. An in-memory B+Tree-based index is utilized to index existingEthereum blocks. The system uses Hadoop to fetch missingblocks from Ethereum clients. We conducted a systematicperformance study, which suggests that the system's per-formance can improve by adding more Hadoop nodes andmore synchronized Ethereum clients.