Scrybe: A Blockchain Ledger for Clinical Trials

Richard R. Brooks, K.C. Wang,
Lu Yu and Jon Oakley
Department of Electrical and
Computer Engineering
Clemson University
Clemson, SC
rrb,kwang,lyu,joakley@g.clemson.edu

Anthony Skjellum
SimCenter &
Dept. of Computer Science and Engineering
University of Tennessee
at Chattanooga
Chattanooga, USA
tony-skjellum@utc.edu

Jihad S. Obeid and Leslie Lenert
Biomedical Informatics Center
Medical University of South Carolina
Charleston, SC
jobeid,lenert@musc.edu

Carl Worley
Department of Computer Science and
Software Engineering
Auburn University, Auburn, AL
crw0034@tigermail.auburn.edu

***Abstract*—The recent popularity of cryptocurrencies has highlighted the versatility and applications of a decentralized, public blockchain. Blockchains provide a data structure that can guarantee both the integrity and non-repudiation of data, as well as providing provenance pertaining to such data. Our novel Lightweight Mining (LWM) algorithm provides these guarantees with minimal resource requirements. Our approach to blockchain-based data provenance, paired with the LWM algorithm, provides the legal and ethical framework for auditors to validate clinical trials, expediting the research process, and saving lives and money in the process.**

Contributions of this paper include the following: we explain how to adapt and apply a novel, blockchain-based provenance system to enhance clinical trial data integrity and non-repudiation. We explain the key features of the Scrybe system that enable this outcome, and we describe resilience of the system to denial of service attacks and repudiation. We conclude that Scrybe can provide a system and method for secure data provenance for clinical trials, consistent with the legal and ethical requirements for the lifecycle management of such data.

I. Introduction

The recent popularity of cryptocurrencies has highlighted the versatility and applications of a decentralized, public blockchain. Bitcoin is the most well-known of such currencies; importantly, the underlying data structure, blockchains, can be applied more broadly than simply to provide a virtual/digital currency. In particular, blockchains provide a data structure that can be used to guarantee integrity and non-repudiation of data, as well as maintaining provenance metadata for systems. Our novel Lightweight Mining (LWM) algorithm provides these guarantees with minimal resource requirements. The LWM algorithm departs from the resource-intensive (and time-consuming) verification approaches that cryptocurrencies, such as Bitcoin, use when expanding the blockchain [1]. Our approach to blockchain-based data provenance, paired with the LWM algorithm, provides the legal and ethical framework for auditors to validate clinical trials, expediting the research process, and saving lives and money in the process.

The remainder of this paper is organized as follows. First, in Section II, we describe the problem statement. Next, in Section III, we summarize *Scrybe*, our novel blockchain-based provenance system. In Section IV, we consider security verification, focusing on data integrity, non-repudiation, and resilience of the Lightweight Mining (LWM) algorithm to distributed denial of service (DDoS) attacks. Finally, we conclude and mention future work in Section V.

II. Problem Statement

Clinical trials test new treatments and pharmaceuticals for countering pathologies. While all science must be performed professionally and be carefully recorded; it is particularly important to track data and documentation changes in clinical trials to satisfy research integrity and regulatory requirements.

Each clinical trial requires a detailed protocol that is approved by the relevant authorities. Once the protocol is in place, an appropriate population of patients is recruited and their consent forms are carefully recorded. A randomization process assigns some patients to a group that uses the proposed treatment, while others are assigned to a control group that receives a placebo.

As the trial progresses, the patient population is treated and raw data is collected. The raw data can take many forms. It may comprise surveys, blood tests, clinical examinations, etc. Either during the treatment or at the end of the trial, the raw data is evaluated to justify the results of the given trial.

It is essential that all information be carefully recorded and not be subject to tampering. Assuring data integrity and veracity enhances the rigor of clinical trials. This ensures the validity of outcomes in compliance with Part 11 of Title 21 of the Code of Federal Regulations required by the United States Food and Drug Administration. Designing trustworthy and reliable systems for managing electronic records along with electronic signatures remains a major challenge in conducting human subjects research.

III. Scrybe: The Blockchain-based Provenance System

This section first overviews Scrybe, our secure provenance system, then considers clinical trials, and implications of integrating and utilizing Scrybe for this application. Subsequently, Section IV provides an explanation of how we verify that Scrybe supports non-repudiation and is also robust against distributed denial of service (DDoS) attacks; in particular, we explain how the Lightweight Mining (LWM) algorithm, a unique feature of Scrybe, is resilient to such attacks.

A. Overview

As illustrated in Figure 1 below, there are two main components of the Scrybe blockchain: blocks and transactions. A blockchain is simply a sequence of blocks, where the current block contains the hash of the previous block.

1) Blocks: As previously mentioned, each block contains the hash of the previous block, which makes the blockchain *immutable*. Blocks are added to the blockchain by *miners*, entities or individuals responsible for maintaining the integrity of the blockchain. Scrybe only allows authorized users to mine blocks through the secure LWM algorithm, which will be discussed in detail in Section IV. Miners are responsible for aggregating a list of transactions and calculating the Merkle root. The Merkle root allows other miners quickly to verify that every transaction is actually included in the block. When a miner is selected to add a block to the blockchain, the block is broadcast to all the other miners, and the data are verified (previous hash, Merkle root, and the miner's signature). At this stage, other miners will be able to detect: if a transaction is omitted from the block, if an unauthorized miner broadcasts a block, and if the miner's signature is invalid.

2) Transactions: Transactions are the backbone of provenance. Transactions can reference previous transactions, providing a chain of custody, or they can be *genesis events*, which register the acquisition of new data. References to other transactions use the *input* fields, while genesis events use *output* fields. The output fields contain persistent URLs (PURLs) that point to the data, along with the SHA-3 hash of the data, ensuring its validity. Additionally, the output fields contain PURLs that point to XML provenance of the clinical trial, along with the SHA-3 hash of the XML provenance.

By storing the SHA-3 hash of the transaction instead of the original transaction, we can drastically reduce the size of the blockchain, and there will be no penalty for an extensive number of inputs and outputs in any given transaction. The original transaction will be stored on a *transaction server*, which will be locally maintained, along with the *data server* and the *metadata server*.

3) Lightweight Mining: Scrybe introduces a novel way to mine new blocks in the blockchain, which is not a difficult proof-of-work required in cryptocurrency applications. The lightweight mining algorithm (LWM) introduced in Scrybe is presented in the following frame.

Lightweight Mining Algorithm (LWM)

Input: The number of miners N.

Algorithm: For each miner $m_i, 1 \leq i \leq N$,

- *Step 1*: m_i generates a random number r_i;
- *Step 2*: m_i broadcasts the SHA-3 hash of the r_i, denoted by $H(r_i)$;
- *Step 3*: Once m_i has collected all N hashes $\{H(r_1), H(r_2), \cdots, H(r_N)\}$, m_i broadcasts the random number r_i.
- *Step 4*: Once m_i has collected all N random numbers $\{r_1, r_2, \cdots, r_N\}$, m_i calculates $m_l = \sum_i r_i \mod N$.
- *Step 5*: m_l is the selected miner to create the next block from the collected transactions.

The purpose of LWM is to provide randomization in miner selection. In a Denial of Service (DoS) attack against Scrybe, we assume a malicious miner targets a particular user by excluding the victim's transactions from the block he or she creates. The randomization offered by LWM guarantees that the victim's transactions will always be integrated sooner or later, as long as there is at least one honest miner.

The core idea of LWM is "sharing-hash-first." If every miner only sends out the random number without sharing the hashes first, a miner can hold his/her own number until he or she has received everyone else's random number. This allows a malicious miner to manipulate the miner-selection by choosing a number that produces a m_l that is in favor of a particular miner or deliberately excludes a particular miner. "Sharing-hash-first" ensures that every miner has to share his/her own number (in the form of hash) with others before they see others' choices. Since hash values are considered impossible to invert in practice, a miner cannot change the random number after the fact. Thus, LWM can tolerate up to $N-1$ malicious miners who collude. As long as there is one miner generating a random number, the modulo operation is randomized.

4) Servers: These locally maintained servers will hold the raw data comprising the ledgers (the blockchains are held in the ledgers). The integrity of the transaction server can be verified by generating a list of all the transactions on the blockchain and comparing that to all the transactions on the transaction server. If there is any discrepancy, the transaction server is deemed disreputable. The integrity of the data and metadata can be verified by comparing the SHA-3 hash of the data to the SHA-3 hash stored in the transaction—if these hashes differ, the relevant server is considered disreputable. The method for storing data on these servers is configurable, and left to the end-user's discretion.

B. Clinical Trials

For clinical trials, first we consider registration of data, then we cover data auditing.

1) Registering Data: Let us consider the case where Alice volunteers as a subject in a cutting-edge clinical research trial run by Bob at the Medical University of South Carolina (MUSC). Once Alice is accepted, her family history is taken, blood is drawn, and vital information recorded. All of this information is collected by Eve and stored on secure local servers. Now, a blockchain transaction can be created. This transaction references the trial genesis event as the sole input and contains a single output. The output consists of a PURL to the results of the blood tests and vitals on MUSC's secure local servers and a hash of those results. The output also contains a PURL to the metadata (time data were collected, family history, and Eve) on MUSC's secure local servers, as well as a hash of the metadata.

This transaction is signed by Bob, and submitted to MUSC's miner. On the miner, Bob's transaction is combined with similar transactions to be included in the next block. Once a miner is chosen by the LWM algorithm, the block containing the hash of Bob's transaction is signed by the authorized miner and added to the local blockchain. The authorized miner then broadcasts the new block to the other miners.

2) Auditing Data: Once Bob has collected sufficient data, auditors from the Food and Drug Administration will review his results. The auditors scan the blockchain all of the transactions pertaining to Bob's clinical trial.

Once the auditors receive authorization from the institution that securely maintains the data locally,they will be able to view the data and all the provenance relating to the trial. The auditors will be able to track subjects, personnel, methods, and results. Once the results are verified, there is a certifiable chain of custody for the trial and all the participants.

IV. Security Verification

Since we replace the resource-intensive Proof of Work (PoW) mining with LWM, a comprehensive security analysis of Scrybe needed to be conducted. Our analysis shows that Scrybe provides data integrity, non-repudiation, and, more importantly, strong resistance to Distributed Denial of Service (DoS) attacks resulting from insider threats.

A. Data Integrity

Integrity means that the corruption of the stored raw data can always be detected. Scrybe uses digital signatures to ensure data integrity. All transactions are signed by the relevant users, and all blocks are signed by the miner who created the block. When the selected miner receives the transaction that contains the hash of the metadata of the clinical trials, he or she can confirm that the transaction was signed with user's private key and that the message was not tampered with in transition.

B. Non-reputation

In this work [2], we assume that the participants could be dishonest, and that it is in the participants' best interests not to disclose their keys or stray from the protocol rules. Non-repudiation is concerned with protecting one participant against possible being cheating by another. We provide the parties with evidence that certain steps of the protocol have occurred. The evidence used in this work is generated by their private keys, which should not be disclosed with others. It is essential, therefore, that the evidence not be modifiable by the recipient. If we can show that the miner could only have come into possession of a message of a certain form if the client had actually sent an appropriately related message to him or her, then we will have shown that the non-repudiation property holds. We use the private keys to generate digital signatures to ensure that the miner can only come into possession of such a message if the client really had previously signed and sent it to that miner.

C. Lightweight Mining Algorithm (LWM) against Insider DoS Attacks

We outlined the novel LWM process for adding to the blockchains in Section III-A3 above. A key feature of LWM is its robustness to malicious miners and Distributed Denial of Service (DDoS) attacks. In order to provide a satisfactory solution, formal verification for the protocol's behavior under threat is highly desirable. A formal verification of the LWM is given in [3]. A Petri Net [4] is used to model the DoS attack, which is then transformed into a Markov chain [5]. We show that there always exists a path from the state representing the victim's transaction to the state where the victim transaction gets added to the blockchain; this proves that the transaction from a victim node will eventually end up in the system as long as there is at least one honest miner. This means that the system is trustworthy as long as there is an active participant who is not malicious.

V. Conclusion

In this work, we adapted the Scrybe provenance system to secure clinical trial data, which allows auditors to verify the integrity of the stored raw data easily. As a blockchain-based ledger system, data integrity and non-repudiation are guaranteed. We tailored the data structures of the transaction and block so that neither raw data nor the metadata is stored on the blockchain itself. This decision allows Scrybe to be isolated completely from the data server for both the raw data and metadata, both classes of which are considered sensitive; this approach greatly reduces the attack surface of the resulting system.

Scrybe incorporates a Lightweight Mining Algorithm (LWM), which we introduced to replace the resource-intensive Proof-of-Work (PoW) mining used by Bitcoin [1]. This design choice means that Scrybe requires fewer resources in terms of both storage space and computation, compared with other blockchain-based tools. A comprehensive security analysis of Scrybe was conducted. Our analysis showed that Scrybe provides many advantageous security feature, including data integrity, non-repudiation, and strong resistance to DDoS attacks arising from insider threats, all of which make Scrybe a sound solution for securely storing clinical trial data.

In future work, queueing theory will be utilized to verify the scalability of Scrybe and quantify the performance and

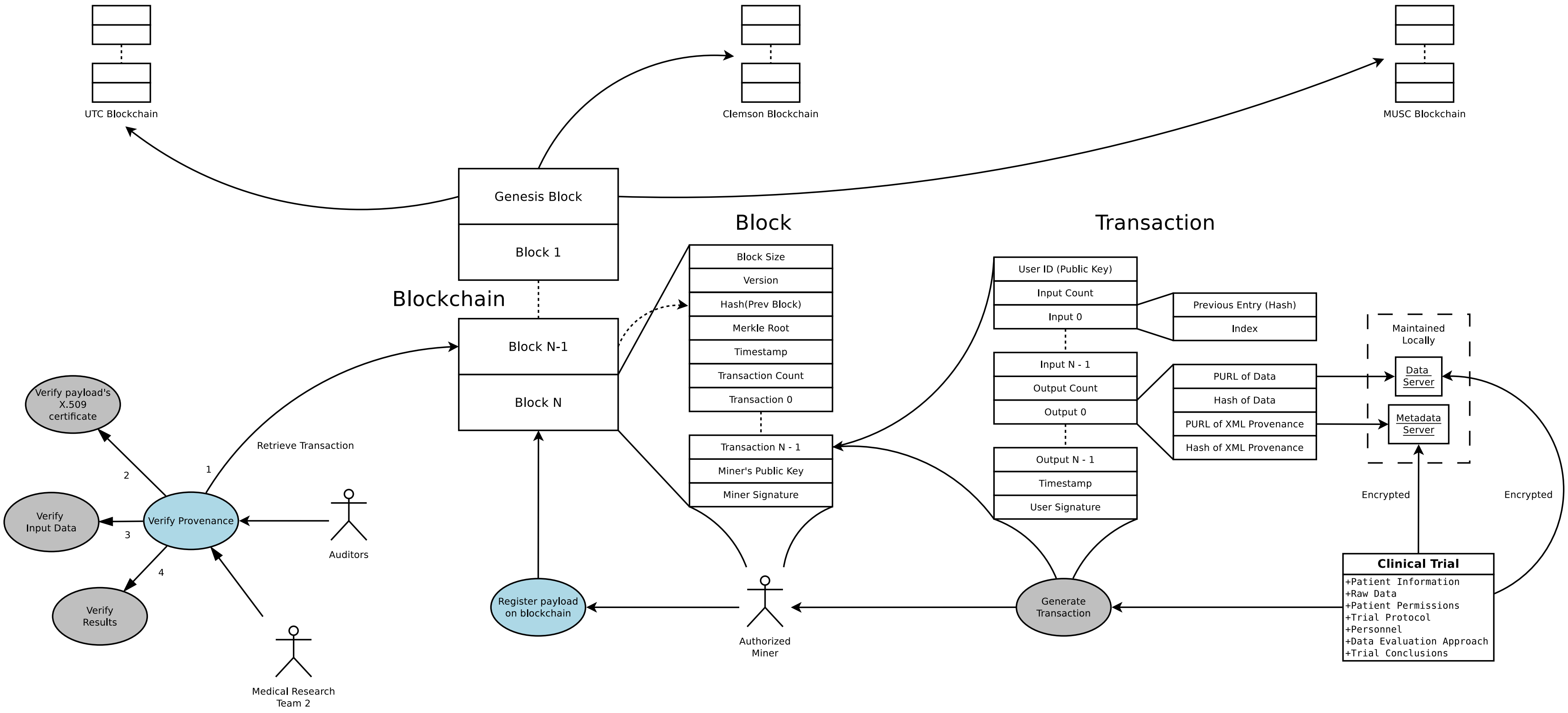

Fig. 1. Scrybe tailored to clinical trial provenance.

efficiency of the system for the clinical trial use case and related applications.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grants Nos. 1547164, 1643020 and 1547245. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] U. Mukhopadhyay, A. Skjellum, O. Hambolu, J. Oakley, L. Yu, and R. Brooks, “A brief survey of cryptocurrency systems,” in *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*. IEEE, 2016, pp. 745–752.

[2] O. Hambolu, L. Yu, J. Oakley, R. R. Brooks, U. Mukhopadhyay, and A. Skjellum, “Provenance threat modeling,” in *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*. IEEE, 2016, pp. 384–387.

[3] O. Hambolu, “Maintaining anonymity and trust,” Ph.D. dissertation, Clemson University, 2018.

[4] R. David and H. Alla, “Petri nets and grafcet: tools for modelling discrete event systems,” 1992.

[5] L. Yu, J. M. Schwier, R. M. Craven, R. R. Brooks, and C. Griffin, “Inferring statistically significant hidden markov models,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1548–1558, 2013.