Use of Blockchain for Monitoring, Identification and Notification of Population Health Trends and Outbreaks

Jesse G. Clark, B.M.

Abstract

The blockchain as a health database could be used to log and identify patient records in a pseudo-anonymous way that would allow researchers to discover real-time public health emergencies and alert professionals and the public. The blockchain would also allow practitioners to digitally sign medical records and documents that could not be altered in the future.

An introduction to the blockchain

The blockchain is a database structure that creates a continuously growing list of blocks that are time-stamped and hold data records that can be written to but not modified[1]. A full copy of the entire blockchain is stored across multiple nodes, or servers, that create a backup as the blockchain is being written.[2]

Each block on the blockchain is made up of a series of transactions. Theses blocks are then distributed from node to node via a peer-to-peer network. When a block is added the nodes verify that the block is valid and that it belongs on the chain[3]. If a block cannot be validated then it is not added to the chain.

The blockchain is the underlying database that Bitcoin uses as a public ledger to keep track of who owns which bitcoins and how much. This paper makes a few references to Bitcoin as it explains public key encryption and the blockchain.

Introduction to public key cryptography

In public key cryptology a pair of keys, comprised of a string of characters, is generated where a public key can be distributed and a private key is kept private. The algorithm that is used to generate the keys is such that just having the public key cannot generate the private key[4].

A message can be encrypted using the public key and only the party with the private key can decrypt the message. A user can utilize the private key to sign a piece of data and prove that they were the original creator. An encrypted message along with the user's public key can digitally sign a message because the only way to create the encrypted message is to do so with the private key. This combination is hashed which creates a digital signature for a piece of data.[5] This hash can be compared against the message and the public key to ensure that the message has not been altered or that the sender is who they claim to be. When this data is added to the blockchain it is time stamped creating a record of its existence.

Using the blockchain to identify users

Bitcoin is pseudo-anonymous in that a user is anonymous until they connect their actions with something that is not anonymous[6]. For example, when a user buys or sells bitcoin with a bank account linked to their information.

For practitioners to use the blockchain in healthcare, they would need to be completely identifiable. When a user writes to the blockchain the data and their hash is signed onto that block. That user could be identified with their public key. However, that user would need to distribute his or her public key in a place where anyone could access it. The public key for a user would need to be verified from the outset that it belongs to that user.

In healthcare the identity of the user would need to be absolutely guaranteed, so a trusted authority would verify each user with their public key[7]. When a health practitioner obtains their professional license it could be linked up with a public key that is on public record. The trusted authority would only need to keep the public key, which would discourage a malicious attack to the system. The trusted authority wouldn't even need to create the private key as the user could generate it themselves and submit the public key when they are verified. The trusted authority would be responsible for linking up public keys with the true identity of an individual.

Securing of private keys

Ensuring that private keys stay private will be one of the biggest challenges of using public key encryption. If someone's private key is compromised, a user could be impersonated and the integrity of the system would be lost[8].

Unlike most passwords, private keys are upwards of a few hundred characters long containing letters, numbers and special characters[9]. These private keys cannot be memorized so the user would need to securely store it.

Bitcoin has introduced a way to create key-pairs off of 12, 18 or 24 four to five-character words which can be more easily memorized by a user[10]. This makes it easier to recover lost key-pairs but comes at a cost. To retrieve the keys the user would need to create the key-pair every time from this list of words. This is cumbersome and could prevent adoption of private key encryption. In addition, many users write the words down which would increase the likelihood of them being found and distributed.

The best way to keep private keys private would be to store the private keys on separate hardware and sign messages offline. This way the users' keys are never exposed to the public. When a practitioner wants to sign a document to be added to the blockchain they would insert their hardware device into a USB port, type in a secure pin and the document could be signed.

If a private key is exposed then procedures must be in place to disable the key-pair. Since the blockchain cannot be altered, that private key would still be used to verify a user's identity up until the point the private key was compromised. At that point the trusted authority would need to revoke the public key and the user would need to be reissued a new key-pair.

If a fraudulent entry was added to the blockchain it could not be removed, rather the timestamp of the entry with the private key could be checked against the trusted authority to determine if the key was valid at that time.

Removing blockchain nodes from the network

Since an individual node can host the entire blockchain, if one node was hit with a denial of service attack that node could be taken offline without causing harm to the network. Since valid key-pairs are required to write to the blockchain an attacker would be blocked when it tried to write its first entry to a node since it wouldn't be valid.

A node could also be taken offline for research. If the data stored on the chain was to be used in a research manner then that single node could be taken off-line while epidemiological data mining was done. After the research concluded that node could join back on the network and would download the missing chain and continue to act as a node.

Using Blockchain as a database

The data on the blockchain is fully accessible to all nodes on the network. This prevents the data from being changed or tampered with. This is ideal for creating a backup of the information but it creates a challenge in limiting read-access to certain parts of data. Data in the blockchain could be encrypted which would allow for an additional level of security, but keys would need to be created and distributed[11].

Encrypting data allows for security of the content but causes two negative issues. The first is that the data cannot be searched easily. When searching for a piece of data that was encrypted, the data would need to be decrypted first before it could be searched. This takes a lot of processing power and time. Second, the encrypted data takes more space and the blockchain is best used with smaller pieces of data.

Storing large amounts of information on the blockchain such as image scans is not currently reasonable because it causes blockchain bloat[12]. The bitcoin blockchain is capped at 1MB per block which limits the number of transactions that can be added to a block[13]. Increasing the block size to a larger amount would allow for the blockchain to accommodate more information. The storage capacity of each node would need to be increased gradually so it could continue to support the size of the blockchain. Although storage space is getting cheaper, those who maintain

blockchain nodes would need to budget for increased storage. This would also create a barrier to entry for new nodes.

To write to the blockchain a user must prove that they are allowed to write. They can do this by signing a message using their private key that would be verified against a trusted authority that would keep a user's public key on file[14]. The host nodes would check the message along with the signature and verify the identity of the user. An attacker trying to write to the blockchain would be blocked because their signature would not match a valid signature held by the trusted authority.

Storing a subset of a patient's health record on the blockchain

One use of the blockchain would be to store patient vitals and lab results that could be used for research and verification. The information could be in shorthand to reduce the size and then extracted and formatted upon retrieval through an API.

The data could also include a way to identify the user utilizing a signature from a key-pair. Similar to a practitioner's signature, a patient could have a key-pair that is used to sign the transaction. The Shared Nationwide Interoperability Roadmap identifies "Accurate Individual Data Matching" as a priority[15]. Each medical record could be signed using the patient's private key and added to the blockchain. Similar to Bitcoin, the user would remain anonymous on the blockchain but could be identified by their practitioner's office. Records could then be linked up or merged using the patient's public key.

Using a second trusted authority, the patient's public key could be on file with demographic information that could be accessed for research purposes. In addition, for practitioners with access, a link to a patient's personally identifiable information could be created. Those without access to the personally identifiable information could still connect multiple health records to a single patient without knowing the patient's identity.

Using a key-pair to sign and identify records goes beyond using a unique identifier as described in the Interoperability Roadmap. With a numeric identifier a record could be altered without anyone knowing. By signing a document with a private key the data can be verified and confirmed that it has not been altered. Adding that data to the blockchain adds a timestamp and further protection that the data cannot be rewritten or deleted.

Blockchain for researchers

As soon as a block is written to the blockchain, researchers would have access to that data and could add it to their data sets. For example, the ability to identify health trends in real-time could speedup identification in the case of an emerging disease outbreak. In an outbreak, the demographics that are on file along with the practitioner that submitted the data could allow for quicker determination of which

geographic areas and populations are most at risk for contracting the disease. In addition practitioners in the area could be alerted to the outbreak status.

If the outbreak affected a specific profile of patients, perhaps those with diabetes, the practitioner could be given a list of patients that match the criteria, identified by their public key. Those patients could be contacted and alerted to specific symptoms that accompany the disease. If early detection was an important factor in preventing disease development, making the public aware of symptoms could greatly decrease the time required for identification.

To expand on this idea, the notification wouldn't need to be done manually. A patient could opt into an automatic notification system that would alert them to the disease outbreak and related symptoms. This notification could be done via email or text message that would not contain any private information. The user would be notified of the outbreak and symptoms and could then login to a secure portal to understand why they received the message (if necessary).

Blockchain for practitioners

As data is written to the blockchain it could be interpreted in real-time and added to existing disease models to better understand patient outcomes and the most effective treatment options.

The practitioner would have access to these models as well as the history of patients with similar symptoms. If the model showed that the majority of patients with those symptoms responded best to a particular treatment method, the practitioner could use that data to inform their treatment plan.

Verifying the authenticity of documents

Storing entire documents on the blockchain may not be feasible but storing the hash of a document to prove its existence would be. The hash of a document, and a document's location, could be stored on the blockchain and linked to a particular practitioner and patient. If the document needed to be verified this could be done by comparing its hash value to the one that is stored on the blockchain. The timestamp on the blockchain would prove its time and the signatures on the document would prove who wrote the document and who the patient was.

Conclusion

Storing population health information on the blockchain would allow for immutable health records that would be pseudo-anonymous using key-pairs. The blockchain would allow for health information to be distributed quickly and enable researchers to mine the data to identify real-time trends. Patients could be notified of emerging public health issues and educated on precautionary measures, alerted to disease symptoms and given direction to seek professional treatment.

Acknowledgements

I would like to thank Michelle Fargher Clark, B.A., M.P.H. candidate, for reviewing and suggesting edits. If you have feedback, please email contact@jessclark.com.

---

[1] "Blockchain (database)". (n.d.). Retrieved August 06, 2016, from https://en.wikipedia.org/wiki/Blockchain_(database)

[2] Trent McConaghy, Rodolphe Marques, Andreas Muller, Dimitri De Jonghe, Troy McConaghy, Greg McMullen, Ryan Henderson, Sylvain Bellemare, and Alberto Granzotto, "BigchainDB: A Scalable Blockchain Database". (June 8, 2016). Retrieved July 22, 2016, from https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf

[3] Watters, Audrey "The Blockchain for Education: An Introduction". (April 7, 2016) Retrieved August 06, 2016, from http://hackeducation.com/2016/04/07/blockchain-education-guide

[4] "Understanding Public Key Cryptology". (May 19, 20105) Retrieved August 06, 2016, from https://technet.microsoft.com/en-us/library/aa998077(v=exchg.65).aspx

[5] CGI Group Inc., "Public Key Encryption and Digital Signature: How do they work?". (2004) Retrieved August 06, 2016 from http://www.cgi.com/files/white-papers/cgi_whpr_35_pki_e.pdf

[6] Andeessen, Marc, "Why Bitcoin Matters". (January 21, 2014) Retrieved August 06, 2016 from http://dealbook.nytimes.com/2014/01/21/why-bitcoin-matters/

[7] AviD, In response to "How should I distribute my public key?". (November 17, 2010) Retrieved August 06, 2016 from http://security.stackexchange.com/questions/406/how-should-i-distribute-my-public-key

[8] Secure Storage of Private Keys. (n.d.) https://technet.microsoft.com/en-us/library/cc962023.aspx

[9] ssh-keygen -t dsa

[10] Marek Palatinus, Pavol Rusnak, Aaron Voisine, Sean Bowe (September 10, 2013) Retrieved August 06, 2016 from https://github.com/bitcoin/bips/blob/master/bip-0039.mediawiki

[11] Lewis, Antony, "So, You Want to Use a Blockchain for That?" (July, 22, 2016) Retrieved August 06, 2016, from http://www.coindesk.com/want-use-blockchain/

[12] Wagner, Andrew, "Ensuring Network Scalability: How to Fight Blockchain Bloat" (November 06, 2014) Retrieved July 22, 2016 from https://bitcoinmagazine.com/articles/how-to-ensure-network-scalibility-fighting-blockchain-bloat-1415304056

[13] Average Block Size, Retrieved August 07, 2016 from https://blockchain.info/charts/avg-block-size

[14] Apodaca, Richard L, "Six Things Bitcoin Users Should Know about Private Keys" (April 23, 2014) http://bitzuma.com/posts/six-things-bitcoin-users-should-know-about-private-keys/

[15] "Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap" (2016) P.36, Retrieved July 19, 2016 from https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf