

Blockchain-enable Contact Tracing for Preserving User Privacy During COVID-19 Outbreak

Md. Murshedul Arifeen¹, Abdullah Al Mamun²,
M. Shamim Kaiser^{2,*} Mufti Mahmud^{3,*}

¹ Dept. of Information and Communication Technology, Bangladesh University of Professionals, Dhaka – 1216, Bangladesh

^{2,*} Institute of Information Technology, Jahangirnagar University, Savar, 1342 – Dhaka, Bangladesh

^{3,*} Dept. of Computing & Technology, School of Science & Technology, Nottingham Trent University, Nottingham, NG11 8NS, UK

* Co-corresponding authors. Emails: mskaiser@juniv.edu (M.S. Kaiser); mufti.mahmud@ntu.ac.uk, muftimahmud@gmail.com (M. Mahmud)

Abstract

Contact tracing has become an indispensable tool of various extensive measures to control the spread of COVID-19 pandemic due to novel coronavirus. This essential tool helps to identify, isolate and quarantine the contacted persons of a COVID-19 patient. However, the existing contact tracing applications developed by various countries, health organizations to trace down the contacts after identifying a COVID-19 patient suffers from several security and privacy concerns. In this work, we have identified those security and privacy issues of several leading contact tracing applications and proposed a blockchain-based framework to overcome the major security and privacy challenges imposed by the applications. We have discussed the security and privacy measures that are achieved by the proposed framework to show the effectiveness against the security and privacy issues raised by the existing mobile contact tracing applications.

1 Introduction

The COVID-19 pandemic caused by SARS-CoV-2 virus is the current growing global pandemic. This deadliest virus spreads from person to person during close contact through small droplets originated from talking, coughing and sneezing. For controlling the spread of the coronavirus, steps should be taken to obstruct person to person transmission. Preventing contact transmission can ensure that no new cases can be generated from each confirmed case of COVID-19. Contact tracing is a systematic approach that helps to identify, assess and manage those people who have been disclosed to a COVID-19 patient so that further transmission in the community can be prevented [1]. This preventive measure is an essential tool to break the human to human virus transmission chain. Each identified person who has been exposed to the disease will be quarantined for 14 days since last contact.

Various governments are releasing mobile applications to trace contacts of COVID-19 patients. However, the approaches for contact tracing never completes

without considering security and privacy measures of users personal data. Since, user's device communicates data from and to the monitoring server, security and privacy of those data becomes a major concern. General users and COVID-19 patients become nervous about the contact tracing applications which continuously broadcast their personal data. Moreover in the context of pandemic situation like COVID-19, data privacy remains an important issue. The user may concern about their collected personal data and raise question regarding who owns their data and how their data will be accessed and manipulated by the monitoring authority. Therefore, trust, security and privacy measures must be taken before launching these contact tracing applications. Several security and privacy issues have been raised already regarding current contact tracing methods applied by several countries or agencies. For instance, Israel government tracks the mobile phones data of suspected COVID-19 infected persons, South Korea government stores personal data of known patients in public database, Taiwan's medical agency keeps track of patients travel history, Singapore government uses a contact tracing app to collect user's location data and mobile phone number and China implemented a surveillance system to monitor public movement [2] [3].

Considering public health Bangladesh government also released an app known as CoronatracerBD which requires location of the user. Though these applications provide accurate contact information but these approaches for tracing contacts by collecting user's personal information violate the privacy rules. The above methods for contact tracing system involves a central authority (like Government or medical agencies) which collects, receive and distribute data to all other users. Therefore, there is no privacy from central authority on the private information of the users. A corrupted person in the central authority may manipulate or inject false data to the server. The users need to trust the monitoring authority blindly for sending their data. Also, the central server may be hacked or the data may be manipulated intentionally, in this case the whole system will fail to operate. Moreover, the centralized infrastructure introduces delayed response, single point failure and sometimes scalability issues. Regarding privacy, existing applications collect location, phone number and can trace user's activity through these data. Therefore we need to consider the privacy issues of these contact tracing applications. In [2], the authors mentioned three notions of privacy in the context of contact tracing system. They are – privacy from snoopers or adversary, privacy from contacts and privacy from the central authorities. Existing solutions like TraceTogether app can overcome the first two privacy notions but along with other applications TraceTogether do not consider the privacy and security issues from authority. Also the recently proposed applications do not consider any proactive measures like cryptography techniques for the collected information. Therefore, we need a distributed and proactive approach where the users data can be shared securely (without manipulation or stealing) and anonymously. Also the minimum possible amount of data should be collected for accurate contact tracing. Private data should be avoided from this system for example location info and phone number.

In this article, we have proposed blockchain which is a distributed decentralized and proactive solution for contact tracing applications. Here public blockchain network is used as a distributed public ledger where any user can register, upload and query their contact list in the blockchain network with security and preserving their privacy.

In section 2 basics of contact tracing mechanism and some leading contact tracing applications are discussed with their privacy and security issues. In section 3 the proposed blockchain based framework is explained. Section 4 discusses the security and privacy measures achieved through the proposed blockchain framework and finally section 5 concludes this article.

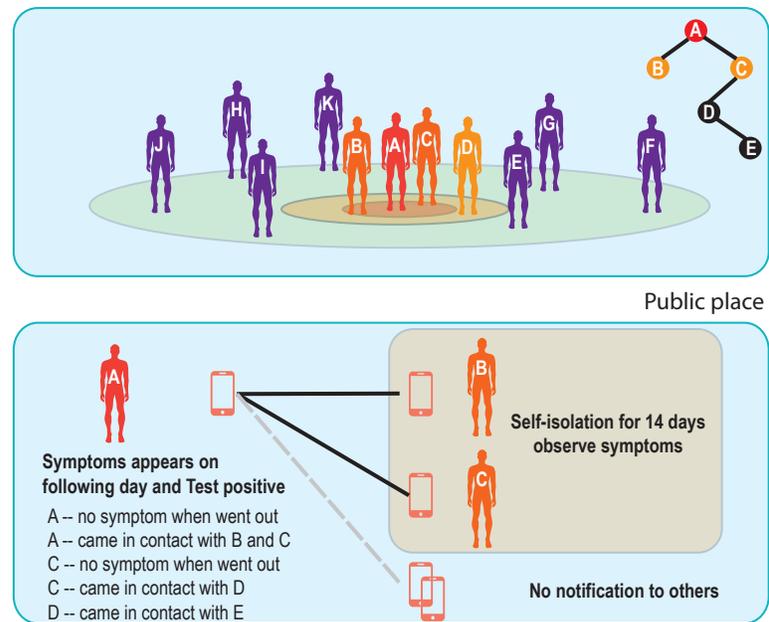


Figure 1. Coronavirus contact tracing app. The app employed Bluetooth interface and trace contact within 2 meter for 15 minutes. A unique ID of each anonymous app user contact tracing data is stored in a server for 14 days. Notification/alert will be sent to the close contacts if symptoms appears in a user and tested positive

2 COVID-19 Contact Tracing

2.1 Contact Tracing Basics

The best way to slow down the spreading of coronavirus from human to human is contact tracing. This process effectively identifies those who are exposed to COVID-19 patients and based on these identifications, necessary steps can be taken like ensuring quarantine. Previously contact tracing was done through manual process like the diagnosed patient is asked with whom he did contact or interact the previous 14 days to identify those people who have been exposed to the disease by this patient. But this manual method of tracing contacts is not efficient. Since, the diagnosed patient may not pay attention to whom he was interacted previously. Also, this slows down the tracing process and introduces human error. To discard the manual process and trace the contacts of COVID-19 patients faster and effectively, we need digital tracing system. Though there are lots of IoT based digital surveillance systems are already deployed around us for various purposes but these systems cannot be utilized as contact tracing system. Since these systems suffer from inherent security and privacy vulnerabilities, so they cannot be used as a health data collecting application in the time of pandemic situation like COVID-19. Currently smartphones are used as a potential tool to implement contact tracing application worldwide. However, smartphone based contact tracing system will be reliable and trustworthy only if the privacy and security of it is ensured.

For contact tracing applications, Bluetooth technology has become one of the popular ways to track the contacts of a user. The basic concept of contact tracing mechanism is as follows- when two user meet but violate the safe distance, their phones application exchange a special packet or key code through Bluetooth signal. When any

user is diagnosed with corona virus then he uploads his coronavirus status and traced contact list in the application database. Distribution of data in the database can be accomplished in two ways- centralized and decentralized way. In the former approach, the mobile phone application uploads its anonymous ID and contact lists collected from other mobile phone users to the centralized database. The server uses the records of contact list in the database to find the contacted user with a COVID-19 patient by analyzing and matching the records. In the later approach, mobile phone uploads only the anonymous ID to the centralized database. The application server then distributes these data to all the users mobile phone. The mobile phone then performs matching and risk analysis operation to alert the user. However, both approach suffers from security and privacy issues due to central database. In figure 1 user A had no symptom when he was outside but then he came in contact with user B and C. Since they violate the safe distance, therefore their contact tracing mobile applications will generate and exchange packets through Bluetooth signal. Similarly user C, D and user D, E will generate and exchange packets. After a period of time, user A tested positive for coronavirus and those who were in contact with user A previously will get notification about the coronavirus status. The notified users will be in self isolation for 14 days and periodically they will be observed for any symptoms. If one of them gets infected then like before his contacted persons will be identified and isolated.

2.2 Existing Contact Tracing Techniques

In this subsection, we have explained some of the leading apps designed worldwide for mitigating the impact of COVID-19. Utilization of these apps help health workers to better understand the behaviour of coronavirus and help to take proper decision. However, these apps show several security and privacy vulnerabilities.

COVID Trace [4] app uses anonymous Bluetooth exposure notification introduced by Apple & Google. It does not collect any personal identifiable data of any user except location and time. When two users are close to each other for 10 minutes, their mobile app exchange and store tokens. When a user tests positive then he or she sends his or her data in the applications server known as escrow. The tokens from the patients made public but this is based on users decision. Other users can download the data and match them with his collected data on his own mobile device. The application tracks previous 3 weeks location data. Location data remains in the user's phone but the application party records location data when the user is outside from home. It reveals that when the application is not tracking that means the user is at home. Based on the users decision, the application party collects approximate area of exposure, lists of symptoms, number of days since first symptoms. The application party anonymizes the location data and uploads it to the server. Also, the time is rounded to nearest hour to preserve privacy. However, this app do not consider any verification whether the user is really infected by COVID-19.

HowWeFeel [5]app collects personal information and other health symptoms for scientists, health professionals and doctors. The aggregated data is securely shared with scientists and doctors who are doing works for COVID-19 to understand the spreading nature of the corona virus, identify the community which may be at severe risk and how steps can be taken to mitigate the impact of COVID-19. The app collects both healthy and infected persons data. The data are collected on the basis of some questions asked by the app for instance how the user feels and whether the user is tested positive for COVID-19 or not. Also unique identifier or token is used to identify each users data uniquely. However, for protecting the users data from the adversaries, the authority uses firewalls, anti-virus software and encryption process to eliminate unauthorized access.

NOVID [6] app uses Bluetooth as well as ultrasonic sound to accurately count the number of contacts and provides precise measurement of distance. Microphones are

Table 1. Existing contact tracing applications

App Name	Data & UID stored and managed by	Types of data collected
COVID Trace	Only location data stored in UE and no UID is used. Application Server collects the data based on users compliance.	Geo-location (exposed area), temporal (number of days since first symptom) and symptoms (all symptoms experienced)
How We Feel	Data and UID are aggregated, stored and managed by the application server. But the data are self reported.	Age, sex, ZIP code and any health symptoms that the user experiences.
NOVID	Random UID and password is used and stored and managed by the application server.	Device model, OS version, language, Bluetooth and sonic signals specification, time and proximity of the contacts
COVID Shield	No data is uploaded into the server and UID is uploaded securely in central server	Random ID, app logs and temporary exposure keys
COVID Near You	Data and UID is stored and managed by applications server	gender, age and IP address
CoEpi	No user UID is used but the collected data is stored in UE	Records of contact list and symptoms list.
COVID Safe Paths	Data is stored and managed by the third party server and UID is stored in UE	Location and time
Safe2	Random Ephemeral UIDs are stored in UE and federated servers are used to handle data	close contacts, location, test results and self-assessment symptoms
Sharetrace	Both are stored in users own server. However, the central server can manage those data	traced contacts, users symptoms and diagnosis results
Zero	Both are stored in users own server. However, the central server can manage those data	Phone number, email address, GPS information and IP address

Legend: UID–Unique Identification number; UE–User Equipment

used to listen only inaudible sounds. With Bluetooth the mobile device sense nearby other mobile devices and microphone detects inaudible sound just passed from other devices. Then it can calculate the proximity between the user and others without the need of location or personal information. When an user registers, a random user ID, password and notification token is generated by the application server for identifying the device uniquely. These information are not linked with other personal information like email, name or device ID until the user permits. The traced contacts alias ID and the contacted users actual ID relation is stored in the server. If an user tests positive then his encrypted user ID is sent to others device and vice versa

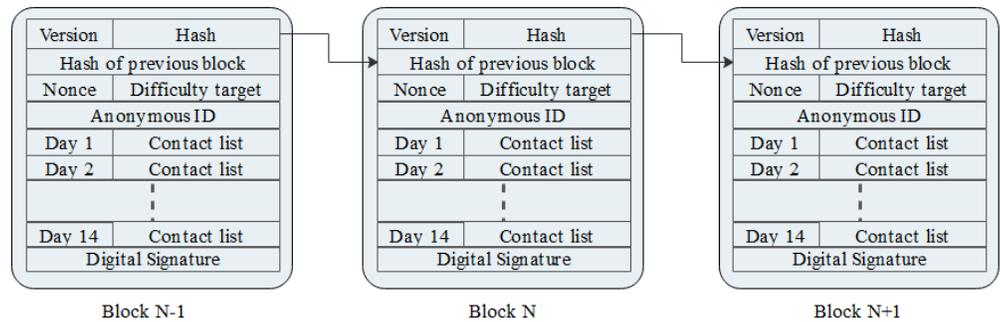


Figure 2. architecture of transaction of blocks. Each transaction contains a 14 days contact list signed by the users private key, an anonymous user ID generated by hashing users public key, hash value of current block and previous block with version, nonce and difficulty target.

COVID Shield [7] app uses Bluetooth technology to collect and share random IDs among the users nearby. If a user diagnosed with COVID-19 then he can anonymously share his random ID for others to know the possible exposure to the disease. This app periodically downloads random IDs from the server and matches them in users own personal device to identify if any possible exposure was occurred in the previous days. COVID Shield comprises of a mobile app, a server and a web portal. No personal data is collected by this application but for exposure notification this app collects random ID, temporary exposure keys and app logs. Google and Apple's Exposure Notification Framework is used to protect the users data. The data are stored in Amazon Web Services which is a third party server. Moreover, the users have complete control over their data collected by the application for example, they can turn off exposure notifications and can delete exposure logs stored in the user equipment. This action will delete all temporary encryption keys stored by the application.

COVID Near you [8] is a reporting website, where users can report their symptoms and testing results. Using these information, the website demonstrates a local and national level illness due to corona virus. Users only upload their symptoms and postal codes. Health officials can view these data but mentioned that they will keep these data confidential and will use for research purpose. SSL encryption is used to protect these data.

CoEpi [9] by default this app stores data on user's device. Like other apps this app also uses Bluetooth to trace contacts. It gives an option to use GPS which shows location but it is based on user's permission. User can edit and review symptoms locally. The server collects data only when the user opts in but the data comprises of anonymized record of contact lists. The server alerts other users to get exposure notification only if the infected user permits. It uses TCN protocol.

COVID SafePaths [10] integrates GPS and Bluetooth with decentralized manner. Users data remains in the personal mobile phone or device but the exposure notification

is extracted from an external source and generated on the users personal device. The app is made with Google Apple Exposure Notification API. Sharing location data with others is optional and depends on the users decision.

Safe2 [11] integrates both Bluetooth and GPS for efficient contact tracing. Bluetooth is used to detect droplet and aerosol (coughing and sneezing) transmission and GPS is used when surface transmission occurs. All data collected by this application remains in users own mobile device until the user gets infected by the corona virus. When infected, the app uploads these data anonymously in exposure alerting system. However, the app authority disclose the data if law enforcement agency requires with court order. SSL encryption technology is used to transfer data from and to the website.

ShareTrace [12] The system produces a random user ID and assigns it to a particular user when he registers to the system. Other information like email, user ID and identifying information remains safe in users PDA. The user ID is used by the PDA to communicate with the ShareTrace server. Each user also has a Bluetooth ID which is used to exchange packets during contact tracing. Each users PDA keeps record of tracing contacts and the users symptoms with diagnosis results. Periodically PDA updates the users information in the ShareTrace server without revealing any identifiable information of users. Data are analyzed by the server.

Zero [13] app collects data and ID. It stores data and ID in the user's mobile device but the exposures event management is performed on a centralized server which is Google Cloud. This application can use both Apple and Google API and TCN protocol. The application party share users personal data with other companies and thrid parties for legitimate business purpose, for law enforcement authorities and several other reasons. The users data may be anonymized for research, public health analyssis and for other legitimate purposes. SSL technique is used to secure the personal data. However, the application party do not take any responsibility if users data is intercepted as mentioned. User may notify the party if they think their data is compromised. The authorities clearly mentioned that they can not ensure complete security of the users data.

All of these applications suffer from some common security and privacy vulnerabilities- Most of them use their own server (central server) for storing and managing the data which leads to single point failure, data manipulation, compromising or stealing. Also, raises trust issues of the users. There is no verification process of users infection status, a user may upload fake data to create panic. In case of NOVID, other users may determine the infected person, if the infected user is the only person they were in close contact in previous week. This is not the case only for NOVID app, other apps also suffer from this issue. Also almost all of the applications collect users personal data which may not convince the user to participate spontaneously shown in table 1 . Besides these apps, other apps also require various types of access from the users personal mobile phone and collects several types of information. For instance, some apps require permission from contacts, photos, media, files, location data, device ID etc and collects users age, email, phone number, IP address etc. Few of them ensure security and privacy but other do not. Since these applications continuously collect and process information including persons private data so ensuring security and privacy is a must before releasing to the general people.

3 Privacy-Preserving Contact Tracing

3.1 Blockchain Network

Blockchain is a hierarchical chain of blocks where a number of transactional records are stored. Each block of the blockchain is secure through a cryptography technique that

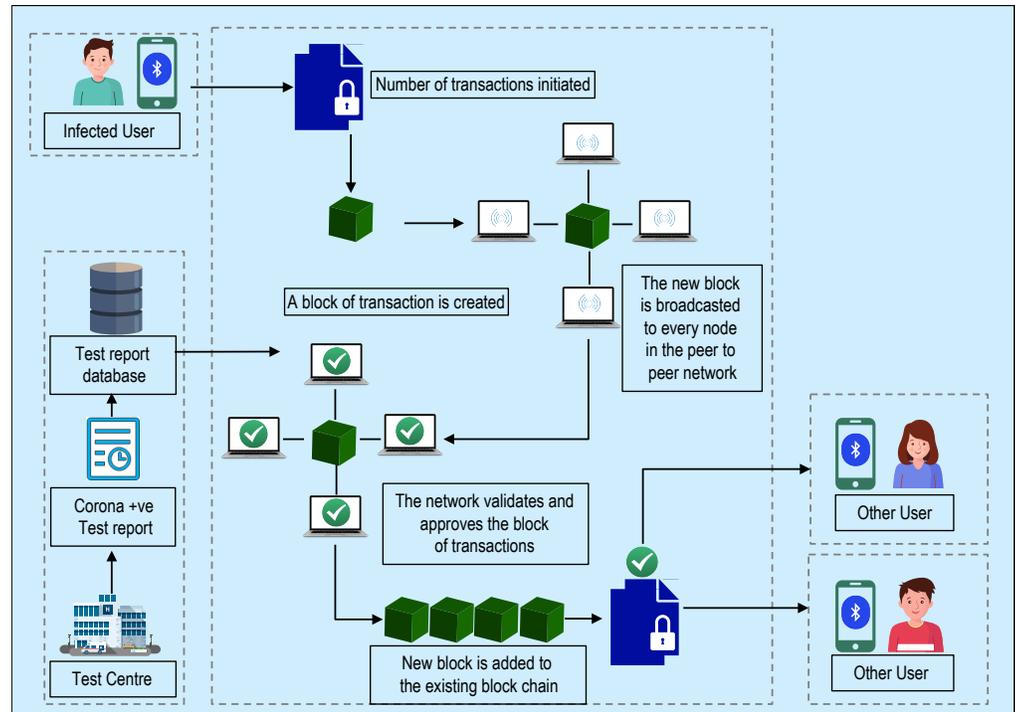


Figure 3. Blockchain based privacy preserving contact tracing framework. Any infected user can upload his contact list through initiating a transaction, the transaction is broadcasted and validated by all the peer nodes with the help of test report database. After that new block is added to the blockchain and other users can get notification or can check their contact status.

ensures integrity of the transaction. Hash value is used to secure each block. Also each block appends the hash value of the previous block which creates a cryptographically secure chain of blocks. Once a block is mined and validated by all nodes in the network and appended to the chain it cannot be altered or manipulated. Therefore the data integrity is maintained and it ensures temper resistance. Thus blockchain can securely stores user data in a distributed fashion omitting any centralized party. So no user need to trust any centralized third party blindly. Since each transaction is signed by the user so authentication is also ensured. Fig 2 shows the structure of the blocks [14]. The block header contains a version number field which indicates software upgrades. Hash field represents the hash value of the current block. Then hash of the previous block field contains hash value of the previous block. Nonce field used for consensus algorithm. Difficulty target field denotes the number of leading zeros. Then starts the body of the block which contains an anonymous ID of the user and last 14 days contact list along with a digital signature.

3.2 Blockchain based contact tracing

A blockchain network (BCN) is a decentralized distributed and secure public ledger that stores records of transactions. In our proposed framework a general user or COVID-19 patient can share his contact list or can know whether he or she was in contact with a COVID-19 patient or not by creating a transaction to the BCN. The transactions will be verified by the mining nodes and broadcasted to all other nodes in BCN. Once every other node verify that the new block of transactions are valid only then the new block

will be added to the blockchain. After successfully publishing the contact lists in the BCN, other users can simply check his corona virus status by querying the blockchain through a query transaction. Here, we did not consider tracing mechanism. We have assumed that any tracing mechanism like TraceTogether which preserves the privacy from contacts and snoopers is the underlying tracing mechanism for exchanging contact packets.

The proposed blockchain based framework is demonstrated in figure 3. For illustrating the mechanism, let's consider, The users mobile application will generate packets like for user A the generated packets are $A = \{A_0, A_1, A_2, \dots\}$ and for user B, $B = \{B_0, B_1, B_2, \dots\}$. When they are in close contact, their contact tracing application will exchange packets and after exchanging, their collected contact list will be $A = \{B_0, E_1, B_2, \dots\}$ and $B = \{A_0, C_1, A_2, \dots\}$. This list of contacts also consists of packets from other users they were exposed to. There will be three types of transactions in the BCN – Access transaction (A_T), Contact transaction (C_T) and Query transaction (Q_T). During registration phase, user A's app will generate a public key, private key pair (various algorithms can be used for example RSA, ECC etc) and a pseudo identity by hashing his public key. Then user A will create an access transaction (A_T) by which he will share his public key and pseudo identity in the BCN. Consider that user B is infected with corona virus and he is now a COVID-19 patient. His corona virus positive certificate is uploaded to the authorized corona positive certificate database by the corresponding health officials and the database is connected to the BCN. Now user B wants to share his contact list of last 14 days in the blockchain network. For this he will create a Contact transaction (C_T) which will contain his last 14 days contact list signed by his private key that is his digital signature will be appended in the transaction along with the header information. The mining nodes will collect this type of transaction and verify it. We did not consider any consensus algorithm because it is time consuming and requires huge resource, since our network do not involves any financial issues or any private information of users. The mining node will simply verify the transaction with the help of the corona positive certificate database. This certificate database is considered since a user may intentionally disseminate false news that he is COVID-19 patient for creating panic. After that this verified block of transactions will be broadcasted to all other nodes in the network. Once all nodes in BCN approve that the block is valid then it will be appended with the blockchain. Now the contact list is visible in the network. Therefore anyone can query the blockchain to know his contact status with any COVID-19 patient. Let's say user D wants to check his status. For this he will create a query transaction (Q_T) to the BC network. Since this transaction requires no validation so a user can simply query the blockchain without verifying his transaction.

4 Discussion

The data originated from the contact tracing mobile applications of a user is important and may contain personal information and should be kept confidential from third parties. Therefore confidentiality of the data should be maintained. Recently proposed centralized server based applications connected with billions of mobile users do not meet this requirement. It suffers from several vulnerabilities and involves single point failure. Also, many to one relationship among the users and servers create network overhead with delayed response [15]. An attacker can compromise the data stored in the server. Blockchain based distributed ledger service can achieve these requirements efficiently and securely. Each block in the chain is secured with its own hash value. So, data integrity is maintained effectively. Also the users send their data through encrypting which keeps them safe from eavesdropper or from other attackers.

The primary goal of contact tracing system is to share one another's data through a

secure and privacy preserving way. A third party server based data distribution mechanism does not provide the security and privacy from contacts. Though the underlying packet exchanging mechanism of TraceTogether application meets this criteria but blockchain also ensures a suitable and trustworthy data sharing process. Moreover, we have used pseudonym based on public key hashing which certainly preserves the privacy of the users.

Since this demonstrated mechanism do not collect users personal information like location, phone number or identity, so this system of tracing contacts is reliable and trustworthy.

5 Conclusion

Personal information in third party server are prone to severe cyber attacks and illegal use. Users should not trust the third party server blindly for their personal and health related data. Also, security and privacy issues of contact tracing apps in the time of pandemic situation like COVID-19 makes users anxious and creates panic which increases unwillingness of downloading and using the contact tracing applications. Therefore, to get a better result from the contact tracing mechanism, we have proposed a secure and privacy aware contact tracing scheme based on blockchain which will inspire the users to use the application. Through blockchain based contact tracing model, users can secure their data and preserve their privacy by monitoring and controlling their data. Users can always check their data uploaded in the blockchain and mitigate the compromising issue by appending digital signature. Also, blockchain ensures that users own their data solely. Moreover, the distributed and decentralized nature of blockchain securely collect, store and share data among all the users. Our proposed blockchain based contact tracing mechanism can also overcome the centralized server problem for contact tracing applications.

References

1. WHO Team. Contact tracing in the context of COVID-19. COVID-19: Surveillance, case investigation and epidemiological protocols. Last accessed: 17th July 2020. Available from: <https://www.who.int/publications/i/item/contact-tracing-in-the-context-of-covid-19>.
2. Hyunghoon Cho, et al. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. 2020. Available from: <https://arxiv.org/abs/2003.11511>
3. Pai Chet Ng et al. COVID-19 and Your Smartphone: BLE-based Smart Contact Tracing. 2020. Available from: <https://arxiv.org/abs/2005.13754>
4. Privacy Policy. COVID Trace. Last accessed: 17th July 2020. Available from: <https://covidtrace.com/privacy-policy/>
5. How We Feel. Last accessed: 17th July 2020. Available from: <https://howwefeel.org/>
6. PRIVACY POLICY. NOVID. Last accessed: 17th July 2020. Available from: <https://www.novid.org/privacy>
7. COVID Shield Privacy policy Last accessed: 17th July 2020. Available from: <https://www.covidshield.app/privacy/>

8. Privacy Policy. Covid Near You. Last accessed: 17th July 2020. Available from: <https://www.covidnearyou.org/ca/en-CA/privacy/>
9. CoEpi: Community Epidemiology in Action. Understanding CoEpi's privacy model. Last accessed: 17th July 2020. Available from: <https://www.coepi.org/privacy/>
10. Contain COVID-19 and Restart the Economy Without Sacrificing Privacy. COVID Safe Paths. Last accessed: 17th July 2020. Available from: <https://pathcheck.org/>
11. Safe2 Privacy Policy. Safe2. Last accessed: 17th July 2020. Available from: <https://safe2.org/privacy-policy/>
12. Privacy. ShareTrace privacy policy. SHARETRACE. Last accessed: 17th July 2020. Available from: <https://www.sharetrace.org/privacy>
13. Privacy Policy. Zero. Last accessed: 17th July 2020. Available from: <https://www.usezero.org/legal/privacy-policy>
14. Minhaj Ahmad Khan et al. IoT security: Review, blockchain solutions, and open challenges. *Future Generation Computer Systems*.2018; 82:395–411. Publisher: Elsevier. doi: <https://doi.org/10.1016/j.future.2017.11.022>
15. Yong Yu, Yannan Li, Junfeng Tian, Jianwei Liu, Blockchain-based solutions to security and privacy issues in the Internet of Things. *IEEE Wireless Communications*.2018; 25(6):12–18. Publisher: IEEE