CovChain: Blockchain-Enabled Identity Preservation and Anti-Infodemics for COVID-19

Pallav Kumar Deb, Graduate Student Member, IEEE, Anandarup Mukherjee, Graduate Student Member, IEEE, and Sudip Misra, Senior Member, IEEE

Abstract—The intangible nature of the COVID-19 virus and the changing strains have created panic and fear among the people, leading to an increase in discrimination of the COVID-19 positive patients and their families in society. Moreover, the spread of infodemics creates daunting challenges for peaceful livelihood and fighting the pandemic. In this article, we propose CovChain, a blockchain-based Internet of Things (IoT) solution for identity preservation and anti-infodemics. Towards this, we propose storing patient information on a private blockchain and exploiting its immutability feature, which helps combat infodemics. We further propose encrypting the data using a distributed Attribute-Based Encryption (d-ABE) scheme to facilitate restricted access to information based on the clearance level. We also propose reducing the load on the miners by using geographically-aware fog nodes. As data in a blockchain is immutable, we delete the blocks corresponding to recovering patients and store the information in a forked CovChain. The cardinality of the two chains helps in maintaining a global census by the cloud servers. Through system implementations, we present the feasibility of CovChain on resource-constrained devices within tolerable delays of 1 second, upload and download rates of 35 Kbps, and CPU (single-core) and memory utilization under 70%.

Index Terms—Internet of Things, Blockchain, Attribute-based encryption, COVID-19, Identity preservation, Infodemics, Healthcare

I. INTRODUCTION

The COVID-19 virus has spread uncontrollably across countries. The changing strains and its effects have created fear and panic among the people. Such fear compels the society to discriminate against those who have tested positive of the COVID-19 virus. The discrimination also extends to the individual’s family members, making life more challenging in these pandemic times. Such issues mandate the need for hiding the identity of the COVID-19 positive patients while keeping the public aware of the patient’s existence. In other words, society must be aware of the number of proximal COVID-19 positive patients without knowing their identities. Moreover, the public is often a target for false news, which creates further panic. In such cases, news or information from reliable and designated authorities is of paramount importance. With the upcoming concept of smart cities, leveraging IoT-based solutions for fighting the pandemic is necessary [1]. Blockchain is one of the technologies that blends edge computing for securing data and offering emergency response, which is particularly important in healthcare [2], particularly in pandemics.

In this work, we address the issues of hiding COVID-19 positive patient identity while making relevant information available to society/public. With the help of off-the-shelf IoT solutions, we attempt to resolve the aforementioned issues and also attempt to stop the spread of false information and infodemics to contain fear and panic within the public. We propose the use of a private blockchain (pBC)-based solution (CovChain) coupled with location-aware foglets, as shown in Fig. 1. A foglet is a collection of fog nodes (FNs) residing in a common network and working in coordination. In the first step, CovChain involves storing the patient information in a distributed fashion. We propose using two pBC networks – 1) COVID-19 positive patients (\(B_1\) in red) and 2) Recovered COVID-19 patients (\(B_2\) in green). We propose storing the information of each infected patient in a single block in the pBC. Additionally, encrypting healthcare data and restricting its access according to the clearance level and attributes is advantageous, particularly using Attribute-Based Encryptions (ABE) in blockchain-enabled scenarios [3]. Consequently, in Step 2, we account for the decentralized and distributed architecture of blockchain deployments, and before inserting the data into the blocks, we propose encrypting the data using
distributed-ABE (d-ABE). Since access control in d-ABE is relatively straightforward, we use it to control information visibility based on the user type. Since we store information related to only one individual in each block, the number of COVID-19 positive patients equals the number of blocks in the pBC. However, the blocks need modification as the patients recover, which is not possible due to the immutable nature of pBC. In Step 3, we overcome this constraint by deleting the blocks corresponding to recovering patients and forking the pBC network for storing the details of recovered patients. The two blockchains make searching and access control more efficient, without violating the blockchain integrity. It also removes the ambiguity that may persist due to unrelated data on a single chain. Further, as shown in Fig. 1, we propose the use of the cloud servers to collect and converge the data from the pBC in each foglet and maintain a global view of the current scenario in each location. Data scientists may perform analytics on these cloud platforms depending on the information available to them.

Example Scenario: Consider a scenario such as the one in Fig. 1 The foglets contain two pBCs with each block containing encrypted information of patients. Concerned authorities such as the police, hospitals, media, and others get access to the blockchain data based on their clearance level. For instance, the police may have access to the location of the infected patients to ensure that they are in quarantine and abiding by the social distancing norms. The doctors in the hospital may have access to the medical history as the COVID-19 virus is an enabler of chronic diseases, particularly respiratory and cardiovascular ones. The media on the other hand gets access to the location-based cardinality of the blockchain so that they are able to report the severity of the spread in a particular area without broadcasting patient details. Similarly, data census authorities get the numbers by finding the cardinality of each pBC without access to private information.

A. Motivation
The rising number of COVID-19 patients has created panic among the people. The limited knowledge about the characteristics and treatment of the COVID-19 virus often leads to the discrimination of the affected patients. This discrimination also extends to the individual’s family members, which further increases the tension and panic. In such cases, restricting access to information to concerned authorities based on their features/attributes is a possible solution for preserving patients’ identities. Additionally, the spread of false news/infodemics is another factor of increasing panic in society. Maintaining reliable sources and retrieving information from the same is necessary for reducing the spread of such infodemics. These factors motivate us to use adopt off-the-shelf IoT solutions and propose a blockchain-based solution coupled with d-ABE for storing patient information. Blockchain inherently takes care of the security and privacy concerns of the data from malicious users and resource allocation-based solutions are also beneficial. Salient features like distributed architecture and immutability help minimize infodemics, and the d-ABE helps in incorporating granular access control.

B. Contribution
In this work, we propose CovChain, a d-ABE-enabled pBC-based solution for preserving COVID-19 patient identities and minimizing the spread of infodemics. The following highlights our specific set of contributions:

- **Access Control**: We encrypt the data using d-ABE before inserting them into the CovChain for providing access to only relevant information to the concerned authorities.
- **Combat infodemics**: As blocks in a pBC are immutable, we store each individual’s information in a separate block, which helps maintain a reliable census of positive and recovered patients.
- **Load Balance**: We introduce a fog computing-based solution to reduce the load from the cloud. It also avoids bottlenecks at the cloud and helps in maintaining location-aware information, which increases data security.

II. RELATED WORK

In this section, we present some of the existing literature towards securing healthcare data and then distributing information using the features of blockchain in IoT environments.

A. Securing Healthcare Data

Security threats to healthcare data start at the device level before initiating communications. Keeping this in context, Tao et al. [5] proposed a hardware-based secure data collection scheme – SecureData. SecureData works in two stages, which involves 1) light-weight encryption and decryption on field-programmable gate arrays (FPGA) and 2) a secured cipher sharing algorithm. Data from these devices are then shared at the cloud/fog level for storage and analysis. In such devices, the healthcare data from multiple sources need to be aggregated, which opens the scope for threats such as confidentiality disclosure. Tang et al. [10] proposed a healthcare data aggregation scheme while preserving patient privacy. Additionally, the authors proposed the addition of noise to the data for enabling differential privacy. The authors in [6] proposed a Secure and Anonymous Biometric Based User Authentication Scheme (SAB-UAS) to facilitate secure communication among the healthcare devices. Apart from conventional methods, Min et al. [7] proposed a reinforcement learning-based data offloading scheme for ensuring the privacy of a patient’s location and usage pattern. The authors further provided evidence of reduced computation latency and energy consumption. Healthcare data usually needs real-time transmission, which opens the scope for the fog computing paradigm. However, ensuring privacy preservation in the real-time transmission of electronic medical records (EMRs) is challenging. The authors in [11] proposed secure methods by introducing identity managers in the fog nodes for overcoming such issues. Moreover, upcoming communication technologies such as 5G have the potential to facilitate eHealth with reduced latencies and increased security.

B. Blockchain & its Applications

Blockchain offers attractive features like secure, reliable, and robust communication. However, its appropriate usage in
IoT environments is a challenging task. To overcome this issue, Memon et al. [12] grouped applications based on real-time, non-real-time, and delay-tolerant requests and proposed schemes for the relevant position of the blockchain. To secure the contents of the blocks in a blockchain, Lei et al. [13] proposed the Groupchain scheme. Groupchain provides security to the data by employing a leader group that is responsible for committing the blocks. The authors in [8] proposed using smart contracts for device-centric trust and enabling training on local data. While this method increases security, it does not ensure data integrity and immutability. The trusted devices may generate malicious data on being subject to external attacks. The authors in [14] proposed the use of blockchain for facilitating a trusted collaboration mechanism for mobile edge computing environments. Smart contracts such as in [9] help in generating alarms/notifications for mobile users when they come in contact with potentially infected individuals. The authors presented an RFID-based proof of concept. However, the proposed smart contract does not address the issue of data privacy, its integrity, and the spread of infodemics.

C. Synthesis

Research on the spread of COVID-19 is important and the existing literature provides numerous solutions for restricting the spread of the virus and the detection of potential patients using IoT solutions. However, there exists a lacuna in maintaining the privacy of patient data and preventing infodemics. Since patients and their families face discrimination from society on testing positive, personal data should only be available to concerned authorities. However, as people should also be aware of the spread of the virus and the count of positive cases, selective information should be available. In this work, we present CovChain as an attempt to address 1) privacy and 2) infodemics to combat COVID-19 with blockchain-based IoT solutions. We present a concise comparison of the existing schemes with CovChain in Table I.

III. System Model

In this section, we first present the entities present in a d-ABE system and then briefly explain the steps involved. We then discuss the proposed CovChain system for identity preservation and anti-infodemics.

A. Distributed Attribute Based Encryption

Typically, d-ABE [15] consists of three entities – 1) master, attribute authorities (AAs), and users. The roles of each of these entities are:

- **Master**: The master is responsible for distributing the secret keys for the users. It may be noted that the master entities do not participate in the attribute key generation process.
- **Attribute Authorities (AAs)**: The AAs are responsible for creating and distributing both public and private attribute keys among users and participants. They are responsible for identifying the attributes pertaining to each user and then determining the corresponding access level. In a d-ABE, there are multiple AAs in the network, and only one AA is responsible for one attribute. However, in this work, depending on the application and demographics, we assign multiple AAs that may set the same attributes. In this work, we assume that the AAs forward the personalized private attribute keys to the users through secure channels.
- **Users**: The users are responsible for encrypting and decrypting the messages. Decrypting the messages in d-ABE is usually straightforward. The decryption keys will only convert the ciphers to plain text in case the attributes match. On the other hand, the users first formulate the access policy and then encrypt using public keys based on the attributes.

These components/entities exchange information among one another based on explicit pairs of public and private keys concerning each user to exchange attribute-based keys necessary for encryption and decryption of messages. Towards this, each of them operates based on seven modules:

- **Setup**: Creation of public (P) and secret (S) master keys for the users to interact with the masters in the network. The P public key is necessary for all operations, and the master S key is necessary for creating user keys.
- **Create User**: This module takes the P and S keys along with the user’s name as inputs for generating the public P_u and secret user keys S_u. The AAs use the P_u to issue on-demand secret keys. On the other hand, S_u is necessary for decrypting the messages.
- **Create Authority**: The AAs run this module only once during initialization. It takes an identifier I_A and generates a secret authority key (S_A). The S_A is necessary for creating attribute keys that take part in the encryption and decryption of messages.
- **Request Attribute (Public)**: The AAs run this module on receiving requests for attribute keys. In case an AA is responsible for the requested attribute (A), it generates the public attribute key (P_A).
- **Request Attribute (Secret)**: This module is analogous

<table>
<thead>
<tr>
<th>Schemes</th>
<th>Centralized (C)/Decentralized (D)</th>
<th>Data Integrity</th>
<th>Immutability</th>
<th>Privacy</th>
<th>Attribute-Aware</th>
<th>Covid-19</th>
<th>Address Infodemics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tao et al. [5]</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Deebak et al. [6]</td>
<td>C</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Min et al. [7]</td>
<td>C</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Rahman et al. [8]</td>
<td>D</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Garg et al. [9]</td>
<td>D</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CovChain</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
to the $P_A$ generation. On receiving requests for secret attribute keys, in addition to the AA’s permission for $A$, it checks the eligibility of $P_A$ for $A$. In case of success, the AA generates and forwards the secret attribute key ($S_A$).

- **Encryption/Decryption:** Both the modules need an access policy (AP) for its operations. The encryption module takes $P$, the message ($m$), $AP$, and the set of public keys ($F$) to generate the cipher text ($c$). On the other hand, the decryption module takes $P$, $c$, $AP$, $S_A$, and the set of user’s secret attribute keys ($S$) to generate the original message $m$.

In this work, we assign the master and AA roles to the miners of the CovChain pBC. Since miners operate on high configuration devices, they may accommodate the key assignment and its distribution with minimal overhead.

### B. Network Architecture

As shown in Fig. 1, we consider a set of users such as doctors, police, paramedics, data scientists, media, and others. The information on COVID-19 patients (positive/recovered) is sent to the miners (after secured key exchange) by the police or doctors (depending on the organizational setting). We consider the miners in the location-aware fog layer. The fog nodes in the same network work in coordination and form foglets. The users interact with the miners in the foglets using radio access network (RAN) technologies using one-hop/multi-hop communication. Depending on the deployment policies, the fog nodes in the foglets may communicate using similar RAN or may have a strong backhaul network. The miners (one or more) in each foglet verify the information’s authenticity and add the information on the CovChain. The data in the chain are encrypted using d-ABE, which assures restricted access to the other participants. For instance, the media and data scientists do not need to know the identity of the patients. They may have access to symptoms, location, and others. The foglets operating on location-centric information reduces the load in the miners. As the miners receive information about a patient’s recovery, they delete the block in the CovChain and fork it (Fig. 2). The forked CovChain contains information of the recovered patients. Deleting the blocks is necessary due to the immutable feature of a pBC. The cloud servers may maintain census by counting the number of blocks by combining information from each geographically spread pBC in the foglets.

It may be noted that we maintain a thick relation between the two blockchains (each for the infected and recovered individuals) to not violate the integrity of the blockchain. As the patients recover, we delete a block from the infected chain only when the following conditions are met:

- Relevant details/payload, apart from status, on the new block for the recovered chain matches the previous one.
- The sum of the number of individuals remains the same before and after completion of the process.

Another approach may be to modify the blockchain by simply modifying the previous hash pointers of the relevant blocks instead of deleting the block. This will not require any change to the data in the block as it will completely change the hash. However, the constraint that the number of patients must remain the same stretches in this case too.

### C. Information Flow

At the inception, when a new device joins the network, it obtains the d-ABE keys from the miners/AAAs to establish secured communication (Step 1 in Fig. 2). The authorized parties upload new d-ABE encrypted information to the geographically assigned miners in the foglets (Step 2). The miners verify the information and add it to the CovChain (Step 3). In the case of patient recovery, the miners fork the CovChain to maintain two chains, one for each positive and recovered patient (Step 4). The other users/participants fetch data from the CovChain and only perceive information based on their attributes (Step 5). The cloud servers may also access information from the CovChain (Step 6). However, the information visibility depends on the users requesting the information.

### IV. Performance Evaluation

In this Section, we explain our experimental setup and present our observations on deploying the proposed scheme.

#### A. Experimental Setup

We consider resource-constrained Raspberry Pi (RPi) devices for sending information to the fog layer. It may be noted that although the secured key exchange procedure is beyond the scope of this work, we exchange the d-ABE keys every time a new device enters the network. The RPIs encrypt the information and send it to the miners. We consider an Intel i5 processor to assume the miners’ role and make decisions on adding/removing blocks from the chain. We consider similar RPI devices for assuming the fog node role, and we use Python 3 to realize the proposed CovChain system. To demonstrate the feasibility of the CovChain, we present the data rates (upload and download) and CPU utilization for each of the devices.
B. Results

In this section, we present the results from our experiments on executing the proposed scheme. We used the `iftop` application in Linux systems to capture upload and download rates. On the other hand, we used `top` to capture CPU and memory utilization.

**Upload Rate:** We observe in Fig. 3 (left) that the devices need a maximum of 35 Kbps for sending the encrypted messages to the miners. We observe that all the devices show similar data rates. We attribute this behavior to them being in the same network while performing our experiments. On the other hand, the miners only need 5 Kbps of upload rate for broadcasting its notifications. We comment on our observation that the devices need minimal data rates for realizing the CovChain scheme. However, using single miners may lead to bottlenecks, which necessitates multiple miners.

**Download Rate:** We present the devices’ download rates on receiving broadcast messages from the miner and that of the miner on receiving the encrypted messages from the devices in Fig. 3 (right). We observe that the devices only demonstrate 5 Kbps on average. On the other hand, the miners have a download rate of 35 Kbps, as all the RPi devices send data to the miner simultaneously. As we increase the number of devices, the download rate of the miner will further increase. Again, in such cases, the use of single miners may cause bottlenecks.

It may be noted that the upload and download rates in Fig. 3 (left and right) correspond to one another and affect the rates accordingly.

**CPU Usage:** Fig. 4 depicts the CPU usage in percentage for both miners and the RPi devices. We record the CPU usage while performing the d-ABE encryption in the devices and decryption in the miners along with the networking operations. On average, we observe that the RPi devices use 70% of a single core. We attribute this to the policy setting and the number of corresponding computations necessary for encrypting the message. On the other hand, we observe less than 10% CPU usage in the miner case. It may be noted that we use a high configuration device for the miners (compared to the RPi devices) and hence notice low usage percentages. From the observation in Fig. 4, we comment that the proposed scheme is feasible for implementation in resource-constrained devices.

**Impact of Increasing Number of Users:** As mentioned earlier, we perform our experiments on a collection of RPi devices. In Fig. 5, we present the memory usage (left) and response time (right) with the increasing number of user devices from an arbitrary RPi. As expected, we observe an increasing trend in both cases. This is because, as the number of devices increases, the data content and the number of blocks also increase. The implicit block replication routine in blockchain also increases the memory demand. On average, an RPi uses almost 67% (including background processes) of the primary memory. However, the rate of increase with the user devices is steady. On the other hand, the response time has a linear increase with the number of devices. This is because an increase in the data leads to the traversal of a longer blockchain for every data request. We observe the same trend on limiting the number of blocks to 25 and 50, respectively. However, in each case, on average, the delay is limited to a little over 1 second. We account for these observations and comment that the proposed CovChain is suitable for resource-constrained devices with tolerable delays.

V. Scope and Limitations

We intend CovChain towards combating the COVID-19 pandemic and minimizing the spread of infodemics. However, it may extend to other IoT applications for similar healthcare
applications for hiding patient identities and other details. Apart from healthcare, the proposed scheme may also extend to industrial scenarios for maintaining immutable data and forking when necessary. Sensor-cloud applications may also use CovChain to hide the identities of the sensor owners and immutable data from the sensors. In each of these applications, the cloud servers may access data from the BCs/pBCs for performing analytics and historical analysis on the same.

Since we use pBCs, the size of the chain keeps increasing exponentially as the virus further spreads. However, as the number of fog nodes increases, the memory in these devices accommodates the blocks with ease. Additionally, regular updates from the authorities are important. Automated methods may be adopted in the future to reduce the dependency on human interventions. Apart from these, the blockchain also suffers from its regular challenges of double-spending, majority attack, bloating, and others. It may be noted that we do not address these issues and this work and plan to resolve them in our extended work.

VI. CONCLUSION

In this paper, we proposed a blockchain-based IoT solution – CovChain – for identity preservation and combating infodemics in the current COVID-19 scenario. We proposed storing data of each COVID-19 positive patient in a single block of the CovChain. As the patient’s recover, we proposed deleting the corresponding block and forking the chain to store the same information. The immutable nature of the pBC mandates deletion of the blocks, and forking the chain assures no loss of patient data. The count of the affected and recovered patients helps in maintaining the uniformity of the census throughout. We further proposed encrypting the data before sending as well as inserting them into the blocks to avoid unauthorized access (since data in pBC is available to participants with the same genesis file). We used d-ABE as our encryption scheme to provide selective access according to the granular properties of users. Through experiments, we demonstrated the feasibility of CovChain using resource-constrained devices.

In the future, we plan to extend our work by incorporating multiple miners and optimizing network usage to avoid bottlenecks and packet drops. We also plan to optimize the CovChain to cope with the exponentially increasing size of the chain.

REFERENCES


BIographies

Pallav Kr. Deb is a PhD Research Scholar in the Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, India. He received his M.Tech degree in Information Technology from Tezpur University, India in 2017. Prior to that, he has completed the B. Tech degree in Computer Science from the Gauhati University, India in 2014. The current research interests of Mr. Deb include UAV swarms, THz Communications, Internet of Things, Cloud Computing, Fog Computing, and Wireless Body Area Networks. His detailed profile can be accessed at https://pallvdebd.github.io/.

Anandarup Mukherjee is a Ph.D. Research Scholar in the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur (IIT Kharagpur). He is also the Director and Co-Founder of the IoT startup, SensorDrops Networks Private Limited (http://www.sensordropsnetworks.com). He has also been associated with various national and international funded projects of organizations such as ITRA- Media Lab Asia, ICAR, British Council, and others. His research interests include, but are not limited to, networked robots, unmanned aerial vehicle swarms, Internet of Things, Industry 4.0, 6G and
THz Networks, and enabling deep learning for these platforms for controls and communications. His detailed profile can be accessed at http://www.anandarup.in

Sudip Misra (M’09 — SM’11) is a Professor and Abdul Kalam Technology Innovation National Fellow in the Department of Computer Science and Engineering at the Indian Institute of Technology Kharagpur. He received his Ph.D. degree in Computer Science from Carleton University, in Ottawa, Canada. His current research interests include Wireless Sensor Networks and Internet of Things. Dr. Misra has been serving as the Associate Editor of different journals such as the IEEE Transactions on Mobile Computing, IEEE Transactions on Vehicular Technology, IEEE Transactions on Sustainable Computing, IEEE Network, and IEEE Systems Journal. He is the Fellow of the National Academy of Sciences (NASI), India, the Institution of Engineering and Technology (IET), UK, British Computer Society (BCS), UK, Royal Society of Public Health (RSPH), UK, and the Institution of Electronics and Telecommunications Engineering (IETE), India. Professor Misra is the distinguished lecturer of the IEEE Communications Society. Further details about him are available at http://cse.iitkgp.ac.in/ smisra/.