A Blockchain based Human-to-Infrastructure Contact Tracing Approach for COVID-19

Danxin Wang, Xianhao Chen, Lan Zhang *Member IEEE*, Yuguang Fang *Fellow, IEEE*, and Chuanhe Huang

***Abstract*—In a post-pandemic era with personal precautions and vaccination, the emergence of COVID-19 variants with higher transmissibility and the socio-economic reopening have raised new challenges to existing human-to-human digital contact tracing systems, where privacy, efficiency and energy consumption issues are major concerns. In this paper, we propose a novel blockchain based human-to-infrastructure contact tracing framework for the post-pandemic era. Specifically, our approach collects and records the interaction information between persons and pre-deployed anchor nodes to trace the possible contacts with confirmed patients, so as to capture the indirect contacts and reduces the energy consumption of users. To address the privacy leakage and reliability issues in contact tracing, we introduce a Self-Sovereign Identity (SSI) model-based blockchain which enables users to gain full control of their own identities and eliminate the linkage between the identity and location information in interaction records. To further preserve the privacy of confirmed patients, we introduce the Private Set Intersection Cardinality (PSI-CA) protocol to estimate the risk of infection by only counting the number of encounters between users and confirmed patients. Two self-executed smart contracts are deployed on the SSI blockchain to perform contact tracing, which guarantees the robustness of the system. The performance analysis validates the effectiveness of our approach.**

***Index Terms*—COVID-19, Digital contact tracing, Blockchain, Privacy-preserving, Self-Sovereign Identity (SSI) model.**

I. Introduction

At the beginning of 2020, there is a new form of viral disease, named COVID-19, detected in human beings. The worldwide epidemic outbreak leads to an unprecedented global crisis with enormous social and economic impacts. The WHO declared the COVID-19 outbreak a public health emergency of international concern on January 30, 2020 [1]. It has been reported that a total of more than 200 million confirmed cases, of which more than 4.25 million have died in 227 countries and regions around the world [2]. It is also observed that the means of preventing the spread of the virus have also changed from total lockdown to a step-by-step process of reopening the economy to live with the virus. Vaccination, masks, and social distance become the most effective preventive measures, and digital contact tracing remains the most powerful way to track confirmed patients [3]. However, the pandemic is still out of control, while the increasing needs for social reopening and the appearance of new variants pose additional challenges to contact tracing.

Digital contact tracing provides technological solutions to automate the contact tracing process so as to quickly and reliably identify persons at risk of infection. The ubiquitous and powerful mobile phones, wearable devices and intelligent sensors are considered as ideal devices to perform Human-to-Human (H2H) contact tracing by using Bluetooth signals to detect encounters with confirmed patients or keeping track of their locations (e.g., via GPS, Cellular and WiFi). Recently, numerous digital contact tracing systems and applications are devised and implemented by industry and academia. For example, Singapore [4] and Google Apple joint consortium [5] proposed the Bluetooth-based contact tracing systems by applying centralized and decentralized architectures, respectively, to monitor people's daily interaction information. China Health code system involved QR code and GPS information associated with users to collect users' activities [6]. These H2H contact tracing systems have been shown to be effective in the early stage of full social and economical lockdown with much less population mobility and activity. However, due to the stagnation of the global economy as a result of the lockdown, we have to conditionally reopen necessary sites for economical and social considerations. In the post-pandemic era, the mobility and density of people have increased, especially in densely populated areas, such as shopping malls, airports and railway stations, frequently and continuously human-to-human interaction recording with the surrounding crowd leads to high energy consumption of H2H contact tracing systems. At the same time, the presence of non-pharmaceutical interventions such as wearing masks greatly reduces the risk of direct contact transmission and the emergence of new variants with high transmissibility (e.g., Delta and Lambda variants) increases the risk of indirect contact transmission (individuals appear at the same location after a contamination event [7]), which reduces the efficiency of H2H contact tracing systems. Based on the above observations, we propose a Human-to-Infrastructure (H2I) contact tracing approach by collecting interaction information between persons and pre-deployed anchor nodes to trace the person who may have contact with confirmed patients, which is more suitable to the post-pandemic era.

Privacy preservation is another challenge in the design of contact tracing systems because of the contact information (the

Danxin Wang and Chuanhe Huang are with the School of Computer Science and Hubei LuoJia Laboratory, Wuhan University, Wuhan, 430072, China (e-mail: wangdanxin@whu.edu.cn; huangch@whu.edu.cn).

Xianhao Chen and Yuguang Fang are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: xianhaochen@ufl.edu; fang@ece.ufl.edu) .

Lan Zhang is with the Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI 49931 USA (e-mail: lanzhang@mtu.edu).

Corresponding author: Chuanhe Huang.

location and identity of users) has to be collected, matched and distributed. In centralized contact tracing systems, there is a trusted central server performing core functionalities such as storing interaction information, risk analysis, and notifications for close contacts [8]. The central server has access to all types of data. Therefore, once the server is compromised by a malicious user, it can cause privacy leakage easily. Additionally, people may be concerned that the central server might be a surveillance tool of the government or other organizations. Hence, there are many solutions using decentralized architecture to perform privacy-preserving contact tracing system. These works assume an honest-but-curious server with minimal engagement in the contact tracing process and offload core functionalities to the user side. Meanwhile, these architectures aim to hide the user identities by generating anonymous IDs for the devices to prevent the server form linking IDs to user information. Nevertheless, these solutions often leverage complex cryptographic algorithms, resulting in significant computational overhead and overcomplicated data management. It is also yet to be seen whether the identity and location information of users are truly desensitized. Hence, we propose a Self-Sovereign Identity (SSI) model based blockchain solution to perform H2I contact tracing. Blockchain can enable users and anchor nodes to desensitize the identity and location information. SSI model allows users to fully control and manage their digital identity personally [9]. In our solution, the blockchain-based SSI model empowers users to manage and store the identity information associated with the interaction records personally and breaks the linkage between the identity and interaction record. In addition, the privacy of confirmed patients is also preserved during the contact tracing process to prevent the discrimination against patients. We introduce the Private Set Intersection Cardinality (PSI-CA) protocol [10] to find only the number of encounters with a patient, and without disclosing the identity and location information of all participants, including patients who have contracted the virus.

Considering all the motivations mentioned above, in this paper, we propose a SSI model based blockchain approach to perform privacy-preserving and efficient H2I contact tracing. The contributions of this paper can be summarized as follows.

- In response to the post-pandemic era with effective preventive measures, we propose an H2I contact tracing approach to fight against both direct and indirect contact based COVID-19 transmissions, which persuasively improves the effectiveness of contact tracing.
- To address the privacy leakage and reliability issues, we introduce an SSI model-based blockchain architecture to achieve reliable distributed contact tracking while breaking the linkage between the identity and trajectory information of users.
- To further preserve the privacy of confirmed patients, we introduce the PSI-CA protocol performed on smart contract deployed in blockchain, which only captures the number of encounters between users and confirmed patients to estimate the risk of infection without disclosing detailed identity and location information.

The remainder of this paper is organized as follows. Section II gives a brief overview of the related work. Section III describes a motivating scenario and preliminaries. Section IV illustrates the system model of the proposed contact tracing system. Section V presents the detailed process of our proposed system. Section VI theoretically analyses the performance. Section VII concludes this paper.

II. Related Work

Digital contact tracing has been considered as a powerful tool for curtailing the spread of COVID-19 pandemic significantly [11]. Existing contact tracing systems mainly record H2H interaction information to find out whether there is a contact with confirmed patients and use Bluetooth signals, Wi-Fi or cellular hotspots, GPS signals, cameras, credit card transaction information, ultrasound, etc., to generate location data. In the literature, there are centralized and decentralized digital contact tracing systems. In centralized systems, there is a central semi-trusted authority, such as the Health or government authority. TraceTogether [4] is developed by the Singapore government that uses Bluetooth Low Energy (LE) to discover and record clients nearby, where users need to submit their contact information to a central server to query whether they have been in close contact with any patient. In this scenario, a user needs to keep the device in active broadcast state, which is energy-intensive. Moreover, Bluetooth technology is vulnerable to many security issues, such as threats of eavesdropping, sniffing and jamming. China Health Code System [6] is jointly launched by the Chinese government and Tencent, which enables contact tracing based on GPS location information and QR code. Many countries such as India (Aarogya Setu) [12], Thailand (ThailandPlus) [13], South Korea (Corona 100m) [14] and Poland (ProteGo) [15] have also introduced contact tracing systems using GPS location information. Luo et al. [16] proposed an acoustic signal based privacy-preserving contact tracing system. These centralized systems can provide more accurate and generalized data. However, such centralized architectures may result in direct privacy leakage to a central server. Therefore, many research efforts focus on the decentralized architecture.

In contrast to the centralized counterparts, decentralized contact tracing service provider does not get access to users' sensitive data due to privacy concerns. Google Apple Contact Tracing [5] employed a decentralized privacy preserving contact tracing based on Bluetooth LE, which uses Associated Encrypted meta-data (AEM) method to encrypt the collected user's Bluetooth signal to protect the user's private data and achieve more accurate proximity matching. MIT developed a Private Automated Contact Tracing (PACT) (EAST-COAST) [17] protocol, which make the user generate and store chirp data (including received time and its bluetooth signal) locally and allows the user store extra metadata, such as location information, in its local log file to help determine the places of encounters. This optional metadata can reduce false positives and increase system accuracy by involving more contextual information. Covidsafe [18] was a tracing protocol proposed by researchers from the University of Washington with a

TABLE I
EXISTING CONTACT TRACING SOLUTIONS

Exsiting works	Technique	Architecture	Power Usage	Coverage	Privacy-preserving	Scalability
Singapore TraceTogether	Bluetooth	Centralized	High	Low	No	Poor
China Health Code System	GPS, QR code	Centralized	Low	High	No	High
ACOUSTIC-TURF	Acoustic signal	Centralized	Low	Medium	Yes	Poor
Google/Apple Contact Tracing	Bluetooth	Decentralized	High	Low	Partially	Poor
PACT(EAST-COAST)	Bluetooth	Decentralized	High	Low	Partially	Poor
CovidSafe	Bluetooth	Decentralized	High	Low	Partially	Poor
CovidWatch	Bluetooth	Decentralized	High	Low	Yes	Poor
Safe paths (US)	GPS	Decentralized	Low	High	Partially	High

very similar process of existing decentralised protocols, which adopted a different key-based generation mechanism to generate the pseudorandom ID to save storage. CovidWatch [19] was developed by the researchers from Stanford University and University of Waterloo, which follows the TCN (Temporary Contact Number) Coalition protocol [20] to generate the key-chain. It guarantees that each key generated from the master key corresponds to one unique temporary contact number. SafePaths [21] employed logging of GPS location trajectories to perform contact tracing and let the confirmed patients share their location trails voluntarily for other users to check whether they have an encounter. These existing decentralized architectures require complex cryptographic mechanisms and signature schemes to achieve key and identity management, making the system inflexible and inefficient. Table I summarizes the features of these existing contact tracing applications.

The requirements of collecting, matching and distributing users' trajectories in contact tracing pose a great challenge to users' privacy protection and identity authentication. Users are seriously concerned about the privacy leakage when generating interaction records with strange entities. With its decentralization, transparency, and tamper-proof features, blockchain can play a neutral role in contact tracing systems, bridging the gap between users, patients and authorities, and reducing the sensitivity of user identity and location information. Xu et al. [22] proposed a blockchain-enabled contact tracing scheme, BeepTrace, to solve the critical privacy-preserving issues by desensitizing the user identity and location information. It involves two distributed blockchains to decouple user privacy: one is tracing chain storing the desensitized personal location information, and the other is the notification chain publishing the match results. In [23], Lv et al. presented decentralized blockchain system (Bychain) with a combination of cryptographic techniques to address the data security of contact tracing and location-proof, and combined zero-knowledge proof and key escrow to solve the identity privacy issue in contact tracing system.

Although decentralized contact tracing systems to some extent mitigate the privacy leakage problem caused by centralized servers, there is still a great challenge to prevent privacy disclosure during the trajectory sharing of users. Self-Sovereign Identity (SSI) model is an identity management model, in which each identity is fully owned, controlled, and managed by the entity to which the identity belongs. It provides a great idea to prevent users from being traced during data sharing to privacy information leakage by breaking the association between user identity and data. Toth et al [24] analyzed and validated fourteen attributes of the SSI model and indicated that the application of SSI model significantly reduces impersonation, fraud and breaches which simplifies user access and prevents providers from collecting large amounts of user privacy information. Liang et al. in [25] proposed a blockchain-based personal health data sharing system under the full control of a user to process user data while ensuring data integrity and privacy protection. Kim et al. [26] introduced a novel scheme Self-Sovereign Privacy (SSP) integrating blockchain and MPC (Multi-Party computing) to protect the privacy and integrity of data collected by IoT devices without a single point of failure while minimizing the cryptographic operations performed on IoT devices.

Similarly, the privacy of confirmed patients also have to be considered during the contact tracing process to prevent discrimination and isolation of such patients. Private Set Intersection (PSI) protocol allows one entity to compute the intersection of its set with another entity without learning any information beyond the intersection, which guarantees privacy protection when sharing data through cryptographic schemes [27]. In [10], De Cristofaro et al. designed an Private Set Intersection Cardinality (PSI-CA) protocol to calculate only the size of set intersection to prevent privacy leakage caused by the presence of both semi-honest and malicious adversaries. Sun et al. [28] proposed a protocol based on Private Set Intersection Cardinality (PSI-CA) to perform the privacy-preserving spatiotemporal matching and optimized it by involving the Bloom filter to enable entities in spatiotemporal matching to adjust the accuracy without disclosing too much private information.

III. MOTIVATING SCENARIOS AND PRELIMINARIES

This section introduces a typical motivating scenario and discusses the key concepts in our proposed contact tracing solutions.

A. A Motivating Example

In this section, we pick up a typical scenario to illustrate how to record interactive information in a restaurant in Post-Pandemic Era. Alice is an ordinary person who wants to have a dinner in a restaurant. There are many diners and waiters/waitresses in the restaurant, and some of them have to take off their masks to eat. Alice is willing to record the interactions with them for contact tracing, but she is also cautious about her privacy and the device energy consumption. Therefore, she wishes to only record the interaction information once without much energy consumption and not disclose her identity information. Jane is a waitress in this restaurant

and serves different diners without masking during working hours, which is recorded in the restaurant's employee system. Jane thinks it is necessary to record the interactions due to the high-risk working environment, but it is very costly to record the interactions with all the diners because that is too much and Jane is also worried about her privacy leakage.

Based on the above observation, we aim to implement a privacy-preserving H2I contact tracing solution. Specifically, we have the following design objectives.

- *Efficient and Low energy consumption.* In densely populated areas, the mobility and density of people have increased, frequently and continuously human-to-human interaction recording with the surrounding crowd leads to high energy consumption and inefficiency of H2H contact tracing systems. Hence, the human to infrastructure contact tracing is more suitable.
- *Privacy preservation.* Users should use pseudonyms to record interaction information to prevent being tracked, and fully control their own identity information. The interaction record stored in an anchor node can only be extracted by a user after identity authentication. In the system, it should separate the user's identity and location information to eliminate the linkage between identity and location information in the interaction record. During contact tracing process, the risk of infection is determined only by the number of encounters with confirmed patients, without exposing any specific information about the user and confirmed patient.
- *Scalability.* The system can support large-scale user interaction record and handle a large number of contact tracing requests at the same time.

B. Preliminaries

To solve the problem illustrated in the motivating example, we need to introduce some preliminaries first.

1) Contact Profile (CP): The contact profile of a user is a collection of spatiotemporal interaction record between the user and encountered anchor nodes. User u's CP_u includes location proofs associated with encountered anchor nodes. It is defined as a set of 2-tuples $p_{n,u} = (\lambda, loc_n)$, indicating that user u is at loc_n in epoch λ, where λ and loc_n denote the time epoch index and the corresponding location index, respectively. In our solution, we assume that the time is divided into epochs of equal length, each represented by a globally unique epoch index λ and each anchor node in different time zones can convert its local time into the corresponding epoch index. Location index loc_n refers to the unique label of an anchor node, which is preassigned when the anchor node is registered in our solution. So the location proof of the encounter between user u and anchor node n can be defined as $p_{n,u} = H(\lambda \parallel loc_n)$, $H : \{0,1\}^* \to \mathbb{Z}$ is a hash function to convert an arbitrary bit string into an integer. Different users who encountered the same anchor node n in the same time epoch λ have the same location proof, it is easy to find out whether they appear in the same place at a same time epoch by checking whether the same elements exist in different users' CPs.

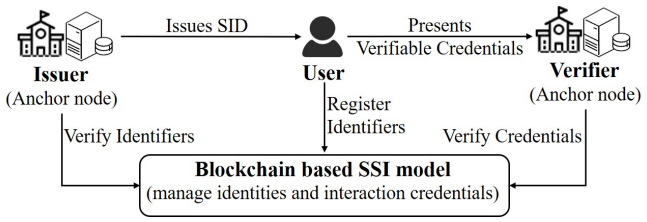

Fig. 1. Self-Sovereign Identity Model

2) PSI-CA protocol based on DDH assumption: Private Set Intersection Cardinality (PSI-CA) [10] is a protocol involving two parties holding sets to compare encrypted versions of these sets in order to compute the cardinality of their set intersection. In our solution, we involve a user with $\mathrm{CP}_u = \{p_{1,u}, \ldots, p_{i,u}\}$, and an confirmed patient with $\mathrm{CP}_{u_{inf}} = \{p_{1,u_{inf}}, \ldots, p_{j,u_{inf}}\}$, yielding the cardinality of the intersection $|I|$, where $I = \mathrm{CP}_u \bigcap \mathrm{CP}_{u_{inf}}$.

Decisional Diffie–Hellman (DDH) assumption [29] is a computational hardness assumption based on the discrete logarithm problem in cyclic groups. The detailed definition of DDH assumption is described as follows.

Definition 1. *Let $\mathbb{G}$ be a cyclic group of order q and with generator g. The DDH assumption states that when g^x and g^y are known, the probability of differentiating g^{xy} from a uniformly random element of $\mathbb{G}$ is negligible (i.e., $\approx$ means approximately equal in probability),*

$$\begin{aligned}(g^x, g^y, g^{xy}) &\approx (g^x, g^y, g^z)\,;\ x, y, z \leftarrow \mathbb{Z}_q \\ &\approx (g^x, g^y, h)\,;\ x, y \leftarrow \mathbb{Z}_q,\ h \leftarrow \mathbb{G}.\end{aligned}$$

In our solution, we exploit the PSI-CA protocol based on DDH assumption to find the cardinality of the intersection of the CPs corresponding to a user and a confirmed patient. Both the user and the confirmed patient need to do some computation on their CP, share the intermediate result to each other through DH approach and then perform the PSI-CA protocol. That is, the user can get the number of encounters with confirmed patients to assess the risk of infection and take measures accordingly. The detailed process of PSI-CA protocol will be presented in Section V.

3) SSI model: Self-Sovereign Identity (SSI) model is a user-centric identity management model, which allows users to fully control and manage their personal digital identity without involving any intermediary. In our solution, we introduce a blockchain-based self-sovereign identity model to enable users to manage and store the identity information associated with the contact records personally. Figure 1 gives the key terms and relationship between the different components in our blockchain-based SSI model. The user is the owner of Sesstional IDentifiers (SIDs) which are issued by the issuer (anchor nodes) according to his/her attributes. The issuers should pre-register in the blockchain and generate verifiable credentials (VCs) associated with the user's ID including his/her signature for verification by others. The verifier receives the credential from the user and verifies it with the system.

Claim is an assertion describing certain attributes associated with the user in typical SSI model. In our solution, we define a Sesstional IDentifier (SID) as the user's identity for each

interaction with anchor nodes, which is issued by the anchor node as a claim.

Verifiable Credential (VC) is a tamper-evident statement associated with user's SID. It can be verified by the public key of the credential issuer without direct involvement of the issuer. A user can have multiple SIDs, each of which can correspond to multiple verifiable credentials at the same time.

IV. System Model

In this section, we introduce the system model and adversary model of our proposed privacy-preserving and efficient H2I contact tracing approach.

A. System Model

The main components of the proposed solution are illustrated in Fig. 2 and are described as follows.

User: The user is the owner of SIDs. A user interacts with different anchor nodes using different SIDs to generate interaction records. A user can have multiple SIDs, each of which also can correspond to multiple interaction records, and there is no connection between SIDs. The user stores SIDs and VCs related to its SIDs locally and submits them to system for contact tracing as needed.

Anchor node: All anchor nodes need to be pre-registered in Health Authority (HA) and form a consortium to jointly manage the SSI blockchain and perform H2I contact tracing by executing smart contracts. Each anchor node has a location index distributed by HA and can change it periodically. The mapping of location index and real physical location is stored and maintained by HA.

During the interaction record collection phase, the anchor node acts as issuers to generate the user's SID and associated VC and store related interaction records locally. There are two types of anchor nodes in our proposed solution. One type of anchor nodes can be some distributed trusted entities such as schools, companies, train stations, etc. They issue long-term effective SIDs and VCs for certain attributes of the user, like the unique number of the student card (which can be used to enter the library and generate interactive records), the unique identity of access card (which can be used to access office buildings and record interaction access), the unique code of ticket (which can record travel information), etc. These long-term effective VCs include the trusted entities' signatures for verification by others and can correspond to multiple interactive records. The other type of anchor nodes are some monitoring facilities deployed by trusted health authority (HA) in densely populated public areas to record the interactions with users. These facilities can detect nearby users' devices though collecting the surrounding environmental signals (Wi-Fi, LTE and Bluetooth signals) which can be considered as a proof associated with a point in space and time [30]. They issue temporary SID and VCs for nearby users, where a credential corresponds to only one interaction record. Similarly, these temporary verifiable interaction attestations also include the facilities' signature for verification by others.

During the contact tracing phase, an anchor node acts as verifiers to verify users' VCs and generate user's CP based on the trust relationship between each other on the signed credentials. In other words, since the long-term effective credentials are issued by trusted entities and the temporary credentials are only issued when the user enters the coverage of an anchor node, so there is an existing trust relationship on each signed credentials. When performing H2I contact tracing, they obtain the user's SIDs and corresponding VCs to generate the user's CP by querying the SSI blockchain and extract interaction records to find the number of infectious contacts (i.e., contacts with confirmed patients).

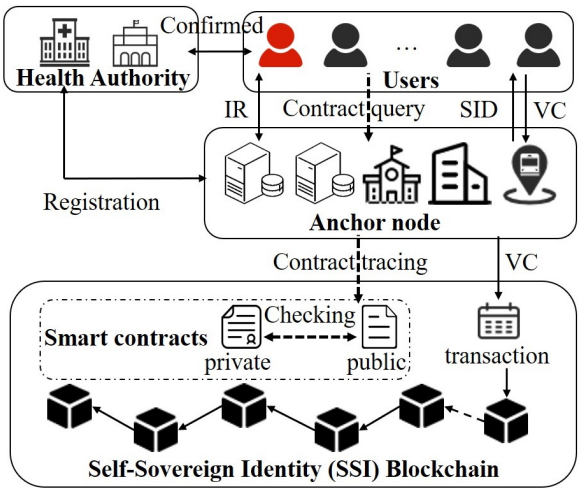

Fig. 2. System model

Health Authority (HA): Health Authority is a trusted third party, e.g., Centers for Disease Control and Prevention (CDC), to manage all anchor nodes (deploy and maintain the monitoring facilities, and supervise distributed trusted entities) and stores the information of the infected patients.

SSI Blockchain: It is a consortium blockchain based on SSI model which is maintained by all anchor nodes for publishing users' VCs. It enables a user to fully control his/her identities related to interaction records and break the linkage between the identity and interaction record. There are two smart contracts deployed on the SSI blockchain to generate users' contract profiles using the extracted interaction records and perform the PSI-CA protocol to find the number of infectious contacts. Any entity can read the information on the blockchain.

B. Adversary Model

We consider two types of attackers in our adversary model and discuss some potential attacks on our proposed contact tracing system.

1) Threat model: We assume that there are two types of attackers in our threat model, malicious users and anchor nodes. A regular user who uses our contact tracing system may reveal their private information, such as identity, location and trajectory, to the system or other entities via interaction recording. The malicious user may discover other users' private information from the public information or compromise the privacy of a confirmed patient from contact tracing result.

We assume that all pre-registered anchor nodes are semi-trusted, which should follow the steps of the protocol honestly but are curious to learn more information from the transcripts

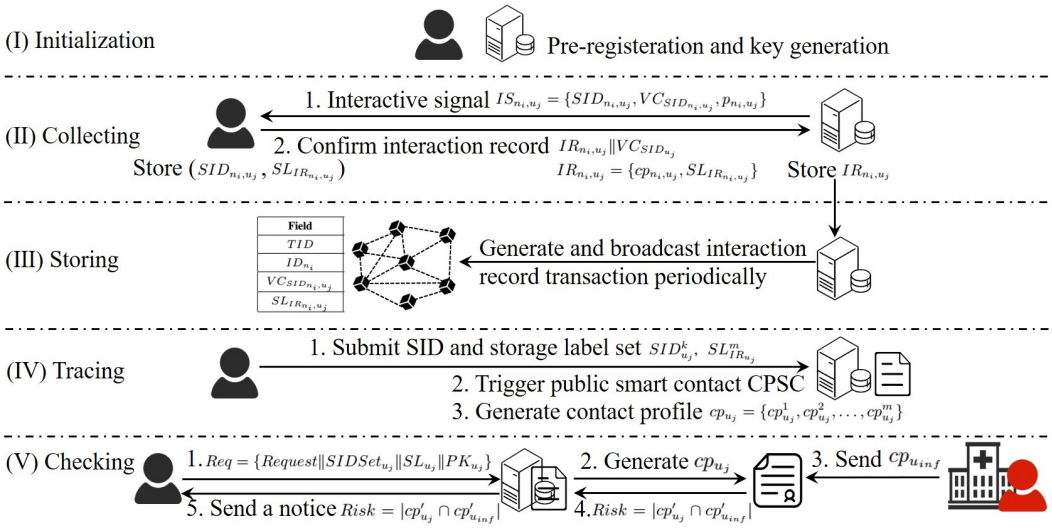

Fig. 3. Overview of proposed system

of received messages. An attacker may compromise the anchor node to inject a fake interaction record or manipulate the records. Compromised anchor nodes may also collude with each other to control the whole blockchain, trace users and manipulate interaction records. We assume that the attacker can only compromise a small part of anchor nodes during any short time period because of overwhelming cost and the real-time supervision of the trusted organizations or health authority.

2) Potential attacks: We list some potential attacks on our proposed contact tracing system.

- *Fake record attack.* A compromised anchor node generates and injects a fake interaction record with users' previous SIDs in order to disturb the contact tracing.
- *Identity forgery attack.* A malicious user generates interaction record using forged identity to pollute the contact profiles of other users.
- *Tracking and deanonymization attack.* A malicious user or anchor node may be interested in exposing the identity and location information to deanonymize the users and generate their social graphs, or inferring the private information of confirmed patients from contact tracing results.
- *Replay attack.* An attacker collects existing broadcast interaction records and then replay it at another time or forward it to a remote anchor node and replay the messages.

V. Our Proposed Contact Tracing Solution

In this section, we describe the details about the proposed blockchain based privacy-preserving H2I contact tracing approach. As shown in Fig.3, the system consists of five phases: system initialization, interaction record collection, interaction records storage in blockchain, contact profile generation, and contact tracing services. The phases and their operations are presented below.

A. System Initialization and Key Generation

In our contact tracing solution, all anchor nodes need to be authorized and identity authenticated by HA (e.g., CDC), in order to become legitimate nodes in SSI blockchain. A legitimate anchor node n_i pre-registers in HA with its identity ID_{n_i} and location and obtains its logic location Loc_{n_i}, public and private keys $(\mathrm{PK}_{n_i}, \mathrm{SK}_{n_i})$ from HA that manages a public key infrastructure (PKI). Each anchor node should publish its identity ID_{n_i} and public key PK_{n_i} in the blockchain and store the mapping of its physical location and logic location Loc_{n_i} in HA. A user u_j also needs to register in the system and generate its public and private keys $(\mathrm{PK}_{u_j}, \mathrm{SK}_{u_j})$ as a legitimate user.

B. Interaction Records Collection

In our proposed H2I contact tracing solution, we collect the interaction data between a user and a fixed anchor node as proof to perform contact tracing. Depending on the two types of anchor nodes, we can record users' interaction data in two ways. The first is that the user interact with anchor nodes (performed as some trusted entities) via long-term effective credentials issued in advance, like swiping the access card or checking tickets. In this case, it is easy for anchor node n_i to generate the interaction record with user u_j using their existing original systems because user u_j has been registered with the anchor node and has pre-issued ID SID_{n_i,u_j} and long-term effective VCs $\mathrm{VC}_{\mathrm{SID}_{n_i,u_j}}$. Anchor node n_i just generates and stores the interaction record $\mathrm{IR}_{n_i,u_j} = \{cp_{n_i,u_j}, \mathrm{SL}_{\mathrm{IS}_{u_j}}\}$ in its local database, where cp_{n_i,u_j} is the location proof with user's security parameter $\alpha, \alpha \leftarrow \mathbb{Z}_q$, $cp_{n_i,u_j} = (p_{n_i,u_j})^\alpha \bmod p$, p and q are two primes (where $q|p-1$). $\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}$ is the storage label of interaction record IR_{n_i,u_j}, which is the signature of the user u_j to SID_{n_i,u_j} and the current timestamp $\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}} = \mathrm{Sig}_{\mathrm{SK}_{u_j}}(\mathrm{Hash}(\mathrm{SID}_{n_i,u_j} \| timestamp))$. Then the anchor node publishes a transaction including the long-term effective VCs $\mathrm{VC}_{\mathrm{SID}_{n_i,u_j}}$ and the storage label $\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}$ into SSI blockchain and sends $\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}$ to u_j for storage.

The second is that the user interacts with anchor node via temporary credentials issued within the coverage of the anchor node. When the anchor node n_i detects a nearby user u_j entering its coverage area, it sends an interactive signal IS_{n_i,u_j},

$$\mathrm{IS}_{n_i,u_j} = \{\mathrm{SID}_{n_i,u_j}, \mathrm{VC}_{\mathrm{SID}_{n_i,u_j}}, p_{n_i,u_j}\},$$

TABLE II
THE STRUCTURE OF INTERACTION RECORD TRANSACTION

Field	Description
TID	Transaction ID
ID_{n_i}	The ID of anchor node
$\mathrm{VC}_{\mathrm{SID}_{n_i,u_j}}$	The verifiable credential of user's SID
$\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}$	The storage label of interaction record

where SID_{n_i,u_j} is the temporary ID of user u_j, $\mathrm{SID}_{n_i,u_j} = \mathrm{Hash}(r_j)$, r_j is a unique identifier of u_j, $\mathrm{VC}_{\mathrm{SID}_{n_i,u_j}}$ is the VCs, which is signed by the private key of the anchor node SK_{n_i}, $\mathrm{VC}_{\mathrm{SID}_{n_i,u_j}} = \mathrm{Sig}_{\mathrm{SK}_{n_i}}(\mathrm{SID}_{n_i,u_j})$, and $p_{n_i,u_j} = H(\lambda \parallel \mathrm{Loc}_{n_i})$ denotes that the user u_j is at location Loc_{n_i} in epoch λ as defined in contact profile (Section III).

When the user u_j receives the interactive signal IS_{n_i,u_j}, it checks whether $\mathrm{VC}_{\mathrm{SID}_{n_i,u_j}}$ is a long-term effective VCs firstly. If not, user u_j stores the temporary ID SID_{n_i,u_j} in local SID set SIDSet_{u_j}. After that, user u_j confirms the interaction record IR_{n_i,u_j},

$$\mathrm{IR}_{n_i,u_j} = \{cp_{n_i,u_j}, \mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}\},$$

similarly cp_{n_i,u_j} is the location proof with user's security parameter $\alpha, \alpha \leftarrow \mathbb{Z}_q$, $cp_{n_i,u_j} = (p_{n_i,u_j})^{\alpha} \bmod p$. $\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}$ is the storage label of interaction record and stored locally with SID_{n_i,u_j} in pairs, where $\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}} = \mathrm{Sig}_{\mathrm{SK}_{u_j}}(\mathrm{Hash}(\mathrm{Hash}(r_i) \| timestamp))$. Then user u_j sends the interaction record $\mathrm{IR}_{n_i,u_j} \| \mathrm{VC}_{\mathrm{SID}_{u_j}}$ to anchor node n_i.

Anchor node n_i stores received interaction records IR_{n_i,u_j} in its local database and publish a transaction including the VCs $\mathrm{VC}_{\mathrm{SID}_{n_i,u_j}}$ and the storage label of interaction record $\mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}$. If a user refuses to participate in H2I contact tracing, the interactive signals can be ignored and the anchor node cannot learn and store any information about him/her. The two types of anchor nodes only lead to different ways of interaction record collection and have no difference in the subsequent contact tracing phase, so we will not distinguish them henceforth.

C. Interaction Record Storage in SSI Blockchain

After each interaction record is stored, anchor node n_i generates interaction record transaction and digitally signs this transaction to guarantee authenticity and accuracy, and then broadcasts it into SSI blockchain. Table II shows the structure of interaction record transaction.

When anchor nodes issue interaction record transactions into the SSI blockchain, anchor nodes collect transactions to generate new blocks by consensus mechanism. The consensus mechanism guarantees the integrity and consistency of the blockchain across distributed anchor nodes, which further determines the performance of the blockchain system. Therefore, depending on application scenarios and performance requirements, such as for network throughput, computing power, storage, and scalability, we can select different consensus mechanisms in contact tracing blockchain, such as Proof-of-Work (PoW), Proof-of-Stake (PoS) and Direct Acyclic Graph (DAG) based consensus mechanisms [31]. We take PoW as an example to explain the detailed consensus mechanism.

Similar to PoW in Bitcoin, the anchor nodes attempt to find their own valid proof-of-work about interaction record transactions (i.e., a hash value meeting a certain level of difficulty, denoted as Difficulty_Level). Each anchor node, regarded as a miner, attempts to find the nonce θ by calculating the hash value of its block based on the previous block hash value, timestamp, the Merkel tree root of the data block and so on (denoted as $\mathrm{Header}_{\mathrm{data}}$). That is to say, the anchor nodes strive to meet $\mathrm{Hash}(\theta + \mathrm{Header}_{\mathrm{data}}) < \mathrm{Difficulty\_Level}$. Difficulty_Level can be adjusted by the system to control the speed of finding the nonce. After finding a valid θ, the fastest anchor node broadcasts the block and θ to other anchor nodes for verifying and auditing. For mutual supervision and verification, other anchor nodes verify the block and broadcast their verification results with their signatures to each other. If other anchor nodes agree on this block, it will be added to the SSI blockchain in a linear, chronological order, and the fastest miner (an anchor node) is rewarded with the corresponding coins. The incentive mechanism depends on the background of the contact tracing system. If the system is a public welfare project, which aims at containing the pandemic, anchor nodes may participate in the process voluntarily for the common good. On the other hand, anchor nodes can obtain certain rewards from consensus process and other query services.

D. Generating User's Contact Profile by CPSC

Each user can submit contact query to the SSI blockchain at any time to generate his/her CP, which can be used to find out the number of contacts with confirmed patients. That is to say, our solution finds the possible encounters between users and confirmed patients by looking for the intersection between their CPs. We design a public smart contract, CPSC, to trace users and generate users' CPs, which runs on the SSI blockchain and is executed and verified by distributed anchor nodes automatically.

When a user submits a contact query to SSI blockchain, the smart contract CPSC is triggered and executed, and a anchor node traces and collects all interaction records for this user to generate his/her CP. The detailed process is as follows. At first, an anchor node obtains user's local SSI set

$$\mathrm{SIDSet}_{u_j} = \{\mathrm{SID}^1_{n_1,u_j}, \mathrm{SID}^2_{n_1,u_j}, \ldots, \mathrm{SID}^k_{n_1,u_j}, \ldots, \mathrm{SID}_{n_i,u_j}\},$$

the corresponding storage labels

$$\{\mathrm{SL}^1_{\mathrm{IR}_{n_1,u_j}}, \mathrm{SL}^2_{\mathrm{IR}_{n_1,u_j}}, \ldots, \mathrm{SL}^m_{\mathrm{IR}_{n_1,u_j}}, \ldots, \mathrm{SL}_{\mathrm{IR}_{n_i,u_j}}\}$$

and public key PK_{u_j} from a user's contact query message. Since a user may interact with the same anchor node generating m temporary SIDs and a long-term effective SID may respond to m interaction records, we simplify the SSI set and storage labels as $\mathrm{SID}^k_{u_j}$ and $\mathrm{SL}^m_{\mathrm{IR}_{u_j}}$, respectively. Then the anchor node searches for the storage label $\mathrm{SL}^m_{\mathrm{IR}_{u_j}}$ in SSI blockchain one by one. When the anchor node finds the VC $\mathrm{VC}^m_{\mathrm{IR}_{u_j}}$ and the corresponding ID_{n_i} of the anchor node that stores the interaction record, it submits $\mathrm{SID}^k_{u_j}$ corresponding to $\mathrm{SL}^m_{\mathrm{IR}_{u_j}}$ to verify its validity via $\mathrm{Ver}(\mathrm{PK}_{\mathrm{ID}_{n_i}}, \mathrm{SID}^k_{u_j}, \mathrm{VC}^m_{\mathrm{IR}_{u_j}})$. If it returns 0, which means this SSI is invalid, the anchor node rejects this request.

Otherwise, the anchor node extracts and sends the location proof $cp^m_{u_j}$. Finally, the anchor node generates the CP_{u_j} for user u_j,

$$\mathrm{CP}_{u_j} = \{cp^1_{u_j}, cp^2_{u_j}, \ldots, cp^m_{u_j}\},$$
$$cp^m_{u_j} = (p^m_{n_i,u_j})^{\alpha} \bmod p.$$

When a user u_{inf} is tested positive for COVID-19, HA requires the patient to submit a contact query to obtain the $\mathrm{CP}_{u_{inf}}$ for confirmed patient u_{inf} by executing the smart contract CPSC,

$$\mathrm{CP}_{u_{inf}} = \{cp^1_{u_{inf}}, cp^2_{u_{inf}}, \ldots, cp^s_{u_{inf}}\},$$
$$cp^s_{u_{inf}} = (p^s_{n_i,u_{inf}})^{\beta} \bmod p.$$

HA stores the $\mathrm{CP}_{u_{inf}}$ and corresponding security parameter $\beta, \beta \leftarrow \mathbb{Z}_q$.

E. Contact Tracing for Users

To prevent attacks that might identify the confirmed patient, our approach uses private set intersection cardinality (PSI-CA) protocol [27] to find the number of infectious contacts. We design a private smart contract, CTSC, to implement the PSI-CA protocol, which is managed by HA and can only be called by the smart contract CPSC. Figure 4 shows the detailed process for contact tracing. Each user u_j in our system can submit contact query Req at any time to check whether he/she has encountered with a confirmed patient,

$$\mathrm{Req} = \{Request \| \mathrm{SIDSet}_{u_j} \| \mathrm{SL}_{u_j} \| \mathrm{PK}_{u_j}\},$$
$$\mathrm{SIDSet}_{u_j} = \{\mathrm{SID}^1_{u_j}, \mathrm{SID}^2_{u_j}, \ldots, \mathrm{SID}^k_{u_j}\},$$
$$\mathrm{SL}_{u_j} = \{\mathrm{SL}^1_{\mathrm{IR}_{u_j}}, \mathrm{SL}^2_{\mathrm{IR}_{u_j}}, \ldots, \mathrm{SL}^m_{\mathrm{IR}_{u_j}}\},$$

where $Request$ contains the date and location range of the user query.

When anchor node n_i receives the contact query request Req, it triggers and executes the smart contract CPSC, then obtains the CP_{u_j} of user u_j.

$$\mathrm{CP}_{u_j} = \{cp^1_{u_j}, cp^2_{u_j}, \ldots, cp^m_{u_j}\}, cp^m_{u_j} = (p^m_{n_i,u_j})^{\alpha} \bmod p.$$

After that, the smart contract CPSC calls the private smart contract CTSC to find the number of encounters with confirmed patients. The detailed process of smart contract CTSC is as follows.

At first, the anchor node sends the $Request$ and CP_{u_j} of user u_j to HA. Then HA queries the $\mathrm{CP}_{u_{inf}}$ and corresponding security parameter β for the eligible confirmed patients,

$$\mathrm{CP}_{u_{inf}} = \{cp^1_{u_{inf}}, cp^2_{u_{inf}}, \ldots, cp^s_{u_{inf}}\},$$
$$cp^s_{u_{inf}} = (p^s_{n_i,u_j})^{\beta} \bmod p.$$

Based on the trust relationship between HA and confirmed patients, it is quite reasonable for confirmed patients to delegate the security parameter β to HA for subsequent calculations in order to reduce consumption. HA calculates CP'_{u_j} with the confirmed patients' security parameter β, shuffles the resulting values and sends CP'_{u_j} and $\mathrm{CP}_{u_{inf}}$ to the anchor node,

$$\mathrm{CP}'_{u_j} = \Pi\{(cp^1_{u_j})^{\beta}, (cp^2_{u_j})^{\beta}, \ldots, (cp^m_{u_j})^{\beta}\},$$

where $\Pi(\cdot)$ is a random permutation function. The anchor node sends CP'_{u_j} and $\mathrm{CP}_{u_{inf}}$ to user u_j.

When the user u_j receives CP'_{u_j} and $\mathrm{CP}_{u_{inf}}$, he/she calculates $\mathrm{CP}'_{u_{inf}}$ with his/her security parameter α first,

$$\mathrm{CP}'_{u_{inf}} = \{(cp^1_{u_{inf}})^{\alpha}, (cp^2_{u_{inf}})^{\alpha}, \ldots, (cp^s_{u_{inf}})^{\alpha}\}.$$

Then the user matches CP'_{u_j} and $\mathrm{CP}'_{u_{inf}}$, and learns the set intersection cardinality $\mathrm{Risk} = |\mathrm{CP}'_{u_j} \cap \mathrm{CP}'_{u_{inf}}|$, which is the number of the infectious encounters.

When the user finds $\mathrm{Risk} \geq 1$, which means he/she has been in contact with some confirmed patients, the user may need to take prompt action, such as COVID-19 testing, self-isolation, etc. If the user is fully vaccinated and has the required precautions in place, such as properly wearing a mask and maintaining social distance, the user can determine if he/she needs to get tested with a small risk value based on his/her physical status. However, we strongly recommend that once $\mathrm{Risk} \geq 1$, he/she should get tested as soon as possible and take precautionary action. Another possible application of our approach is that HA can send alert messages to places where infectious contacts are identified. The business owners can take timely actions such as full disinfection to prevent further spread of the pandemic.

VI. Performance Analysis

In this section, we conduct performance analysis for our proposed H2I contact tracing approach in terms of privacy, security, reliability and scalability. We also analyze the data storage and computational complexity of our solution. Meanwhile, we compare our proposed solution with other existing contact tracing approaches.

A. Analysis

The proposed contact tracing solution involves the collection and use of personal sensitive data, such as their identities, locations, trajectories and infection status. It is evident that the use of the personal sensitive information raises serious concerns for users about system security and personal privacy. Meanwhile, due to the full-scale outbreak and rapid spread of the pandemic, the reliability and scalability of contact tracing system also are of vital importance.

1) Privacy: In our H2I contact tracing solution, we provide strong privacy assurance for all users and anchor nodes. For users, during the record collection and storage processes, the user generates and shares a VC associated with the temporary SID of the interaction, which has been cryptographically hashed and signed by the issuer, and is verifiable and unforgeable. Therefore, the VC is issued and stored on the SSI blockchain without revealing the user's real identity. The user also generates and sends a encrypted location proof of the interaction to anchor node for storage. Due to the nature of DDH assumptions and cryptographic hash function, the anchor node cannot obtain any private information about the user except for recording this interaction. The user's identity credential and location corresponding to the interaction record are stored separately on the user and anchor nodes, which enable the user to fully control his/her identities and break the linkage between

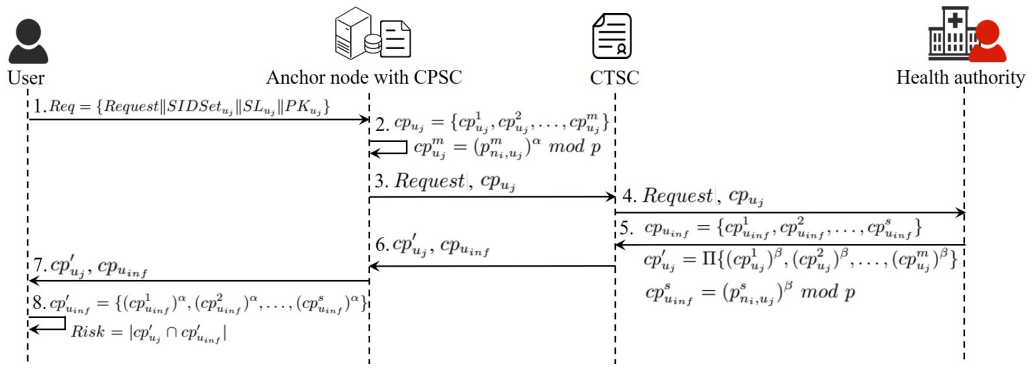

Fig. 4. The process of contact tracing for common users

identity and location, thus preventing the user from being tracked. During the contact tracing process, the generation and matching of user's CPs are performed by two smart contracts CPSC and CTSC deployed on SSI blockchain, that are completely autonomous, self-executed and self-maintained in the form of computer codes. As a result, other users and anchor nodes cannot interfere and manipulate the process of contact tracing and learns any private information during execution of smart contracts. Meanwhile, the introduction of the PSI-CA protocol ensures that any information except the number of encounters cannot be captured by the user during the matching process as proven in [27]. The HA as a trusted organization, can only access the CPs for confirmed patients (which is encrypted by patients themselves) for outbreak control, but nothing about other.

For anchor nodes, they only record the present interaction under the permission of a user, but they are curious about the trajectories of the users as well. Fortunately, in our solution, the location proofs and identities associated with different interaction between the same user and an anchor node are completely different because they are formed based on the user's current spatio-temporal location. The anchor node is unable to track the user through the identity credentials issued on the SSI blockchain. Also it is impossible for the anchor node to access the interaction records stored in other anchor nodes, since they can only be extracted upon user's identity authentication, and the user gains full control of these identities. Furthermore, if a portion of the compromised anchor nodes collude with each other and share their interaction records, they still cannot generate the user's trajectory because they cannot recognize the interaction records from the same user. The anchor node performs contact tracing by executing the smart contracts. It can be aware the location index of the visited anchor nodes when extracting the interaction record, but it is difficult to associate these interaction records with the user's real identity.

2) Security: Next, we analyze the security of our solution. We are particularly concerned about the robustness of our system against the potential attacks listed in adversary model.

Fake record attack. A compromised anchor node may attempt to inject fake interaction records to pollute user's CP. In our solution, the interaction record co-exists with user's identity in pairs. Although the anchor node generates a fake record with users' previous SID and adds it into contact profile successfully, it is still unable to disturb the contact tracing without the location proof which needs encryption via user's security parameter. In some cases, a compromised anchor node attempts to modify the identity information on the blockchain to disrupt a user's CP. However the consistency and transparency of the blockchain ensure that a small number of malicious nodes cannot control and modify the data on the blockchain.

Identity forgery attack. A malicious user may generate interaction record using forged identity to pollute other user's CP. In our solution, a user's identity is issued by anchor nodes, for long-terms effective SID, the forged identity can be easily recognized. For temporary effective SID, it is based on the location tag generated by the spatio-temporal signals of the user's current environment, which cannot be forged.

Tracking and deanonymization attack. A malicious user or anchor node may expose the private information to deanonymize the users and generate their social graphs. However, in our solution, according to the description in the previous section on privacy, a user takes full control of the identity information, while the location information can only be extracted upon the identity authentication, and the linkage between the identity and location information is broken, all of which makes it almost impossible for the anchor nodes and other users to obtain any private information about a user during the contact tracing process. However, if there is a powerful eavesdropper could monitor all nodes in the network and retrieve a user's trajectory by gathering the location indexes of the anchor nodes associated with the interaction records, the attacker may infer some private information of the user (e.g., home address, work unit) in combination with information of compromised anchor nodes and the user's interaction records, which shifts to the area of social engineering. In fact, the attacker can only compromise a small part of anchor nodes during any short time period because of overwhelming cost and the real-time supervision of HA, so sporadic leakage of interaction records is reasonable for the user.

Replay attack. A attacker may collect existing broadcast interaction records and then replay the records at another time or forward them to a remote anchor node for replay. In our solution, it is impossible to generate interaction records with another anchor node if only the malicious user replays the previous SID, because the anchor node will verify whether the user is currently in its coverage. If a malicious user colludes

TABLE III
DATA STORAGE OF OUR PROPOSED CONTACT TRACING SOLUTIONS

Entities	Initialization phase	Interaction record phase	Contact query phase
Users	Public and private keys	SIDs, Storage labels of interaction record	Risk
Anchor nodes	Public and private keys	Part of SSI blockchain, Verifiable credentials, Location proofs	Encrypted contact profiles of user
Health Authority	Real location of anchor nodes	-	Encrypted contact profiles of patients

with compromised anchor nodes to replay past records, it also does not affect the result of contact tracing. This is because during the matching process of PSI-CA protocol, the locations will not be matched successfully because the location proofs are not encrypted with the corresponding security parameters.

3) Reliability: Next, we analyze the reliability of our approach. In the post-pandemic era, our proposed H2I contact tracing solution is more suitable to densely populated areas with high mobility, in the sense that it effectively reduces the frequency of interaction recording and energy consumption. At the same time, the presence of basic preventive measures (such as masks, social distance) and the emergence of new variants with higher transmissibility result in a severer threat of indirect transmission, which can be effectively traced through the H2I contact tracing. Anchor nodes in different regions can control the accuracy of contact tracing by adjusting the length of time epoch. For example, in a restaurant, the time epoch can be set to 15 minutes such that most of contacts can be captured. The combination of H2I and H2H can be deployed in high-risk areas to improve the accuracy of contact tracing.

In our solution, the anchor node generates a unique user SID by collecting environmental signals corresponding to the user within its coverage, and the correctness of this representation of spatio-temporal location is verified in [32]. It prevents users from under-reporting or omitting their trajectory in contact tracing. Meanwhile, the introduction of smart contracts with secure self-execution ensures the reliability of our solution.

4) Scalability: One great challenge of our approach comes from the high cost on anchor nodes to maintain the blockchain, especially considering that the number of records may be huge. Actually, the anchor node can adjust the period of issuing transactions and consolidate multiple interaction records into one transaction on SSI blockchain. At the same time, users travelling over long distances also challenges our contact tracing solution, such as cross-country travel, which needs to query remote anchor nodes upon generating CPs. It will reduce the efficiency of our contact tracing system. Fortunately, such scenarios are relatively rare in reality, as well as there are some additional information (such as flight tickets) to assist in the query process to make it easily.

B. Performance evaluation

1) Data storage: At first, we briefly analyze the data storage of our proposed H2I contact tracing approach. Table III gives the data stored at users and anchor nodes at different phases. For users, they just store their SIDs and the storage labels of the interaction records in pairs and submit them for contact tracing. For anchor nodes, we assume that the anchor nodes have sufficient computing and storage capacities to maintain the SSI blockchain and store the interaction records. Due to the heterogeneity in computing and storage capabilities, some anchor nodes are allowed to only store part of the blockchain. However, they should at least have the local storage space to store interaction records.

TABLE IV
COMPUTATIONAL COMPLEXITY OF OUR PROPOSED SOLUTION

	Computation Complexity	
	Users	**Anchor nodes**
Interaction record	$O(n^2(\log n + 11))$	$O(k^2(\log k + 11))$
Contact profile	-	$O(mk^2(\log k + 11))$
PSI-CA	users: $O(\tau(m\theta^2 + s\theta^2) + ms)$ patients: $O(\tau s\theta^2))$	$O(\tau m\theta^2)$

2) Computational complexity: Next, we analyze the computational complexity of our H2I contact tracing solution. Table IV shows the computational complexity at different phases. During the process of generating and collecting interaction records, a user is required to sign the storage label of the interaction record. Taking the Elliptic Curve Digital Signature Algorithm (ECDSA) and SHA-256 as an example, the computational complexity of the signing process is $O(n^2(\log n{+}11))$, where n is the input size in bits. The anchor nodes are required to sign SIDs as verifiable credentials, where the computational complexity of the signing process is $O(k^2(\log\ k + 11))$ with k being the input size (in bits). Since the generated location proofs during this process are used in the PSI-CA protocol, we treat them as a part of the PSI-CA protocol to compute the complexity. During the process of contact profile generation, a user just submits a query request, so we can ignore the computation overhead. A anchor node has to verify the signature of the user's SID to extract the corresponding interaction record, and therefore the computational complexity of the verifying process is $O(mk^2(\log\ k + 11))$, where m is the number of elements in a contact profile. In fact, there are many existing researches to improve the performance of ECDSA, such as removing modular inversion to reduce its complexity, which is beyond the scope of this paper and will not be explained in detail.

The computational complexity of the contact tracing phase mainly comes from the PSI-CA protocol. For users, the complexity of location proof upon interaction record generation is one modular exponentiation with computational complexity $O(\tau\theta^2))$, which depends on the size of the base θ (the size of p_{n_i,u_j} in bits) and exponent τ (the size of security parameter α or β in bits). When a user submits a Req to perform PSI-CA protocol, the computational complexity is $O(\tau(m\theta^2 + s\theta^2) + ms)$, m, s are the sizes of user's and patients' CPs, respectively. $O(ms)$ is the computational complexity of searching for the intersection. The confirmed

TABLE V
COMPARISON WITH EXISTING BLOCKCHAIN-BASED CONTACT TRACING SYSTEMS

Solutions	Communication	Architecture	Technique	Blockchain role	Smart contract	Description
BeepTrace [22]	Bluetooth, WiFi, Celluar, GPS	Permissioned Blockchain	Public Key Infrastructure (PKI)	Two blockchain for tracing and notification	No	A hybrid system with servers and blockchains.
Bychain [23]	GPS, Short-Range Communication (SRC)	Permissionless Blockchain	Crucial escrow and zero-knowledge proof	Proof of Location for SRC Record Generation, Distributed database	Yes	A novel location based consensus algorithm.
Our proposed solution	Bluetooth, Wi-Fi, LTE	Consortium Blockchain	SSI model, PSI-CA protocol	SSI blockchain	Yes	A novel H2I contact tracing approach.

patient is only required to calculate the location proof, and the computational complexity is $O(\tau s\theta^2))$. A anchor node takes the computational overhead of the other processes in the PSI-CA protocol with the computational complexity $O(\tau m\theta^2)$.

3) Comparison with existing contact tracing systems: Finally, we compare our proposed solution with two existing blockchain-based solutions. Table V shows the comparison between these various solutions. As can be seen from this table, there are different approaches and techniques involved in blockchain-based contact tracing systems. Although BeepTrace [22] also desensitizes the user identity and location information in contact tracing system as in our solution, it demands frequent global geodata update and the computing resources required for geodata matching, which poses as a significant challenge to the scalability of the system. Bychain [23] employs complex cryptographic mechanisms (such as key escrow and zero-knowledge proof) to obtain location privacy and security. However, when the IoT witnesses (e.g., BLE beacons) are densely deployed, computational overhead will increase significantly and the storage of user devices may be run out rapidly. In our solution, the design of SSI-model based blockchain ensures the scalability of the system while preventing privacy leakage by eliminating the linkage between user identity and location. The introduction of the PSI-CA protocol also guarantees privacy preservation for both users and confirmed patients in the contact matching process. Considering the growth in sizes of contact records, our system is practical. The interaction records in our proposed system are only related to a user's own trajectory, and will not cause bursty growth as in H2H contact tracing scheme when a large number of people in a close proximity contact with each other for contact records.

VII. CONCLUSION

In this paper, we have proposed a novel blockchain based H2I contact tracing approach to address security, privacy, efficiency and energy consumption issues for digital contact tracing to fight against the COVID-19 pandemic. In our proposed solution, the novel H2I contact tracing architecture is an ingenious way to capture the indirect contact that is pervasive in the post-pandemic era, and reduce energy consumption compared with traditional H2H contact tracing systems. We present an SSI model based blockchain to collect and store the interaction records, which enables a user to fully control their own identity information and eliminate the linkage between identity and location information in interaction record. This makes it almost impossible to track and identify users with data stored on the blockchain. To further protect the privacy of confirmed patients, we employ the PSI-CA protocol to perform contact matching, which only acquires the number of encounters between users and confirmed patients. The contact tracing is performed by two smart contracts deployed on SSI blockchain, which are completely autonomous, self-executed and selfmaintained in the form of computer codes, thereby guaranteeing the robustness of our solution. The performance analysis further validates the effectiveness of our solution.

ACKNOWLEDGMENT

This research was supported in part by the National Natural Science Foundation of China, No. 61772385. The work of X. Chen and Y. Fang was partially supported by US National Science Foundation under IIS-1722791.

REFERENCES

[1] World Health Organization (WHO), "WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV)," https://covid19.who.int/, Jan 2020.
[2] J. H. C. R. Center, "Covid-19 dashboard," https://coronavirus.jhu.edu/map.html.
[3] A. M. Rahmani and S. Y. H. Mirmahaleh, "Coronavirus disease (covid-19) prevention and treatment methods and effective parameters: A systematic literature review," *Sustainable cities and society*, p. 102568, 2020.
[4] J. Bay, J. Kek, A. Tan, C. S. Hau, L. Yongquan, J. Tan, and T. A. Quy, "Bluetrace: A privacy-preserving protocol for community-driven contact tracing across borders," *Government Technology Agency-Singapore, Tech. Rep*, 2020.
[5] A. Inc. and G. LLC, "Exposure notification," https://www.google.com/covid19/exposurenotifications/, May 2020.
[6] P. Mozur, R. Zhong, and A. Krolik, "In coronavirus fight, china gives citizens a color code, with red flags," *The New York Times*, vol. 1, 2020.
[7] X. Liu, N. Trieu, E. M. Kornaropoulos, and D. Song, "Beetrace: a unified platform for secure contact tracing that breaks data silos," *arXiv preprint arXiv:2007.02285*, 2020.
[8] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha, "A survey of covid-19 contact tracing apps," *IEEE access*, vol. 8, pp. 134 577–134 601, 2020.
[9] A. Mühle, A. Grüner, T. Gayvoronskaya, and C. Meinel, "A survey on essential components of a self-sovereign identity," *Computer Science Review*, vol. 30, pp. 80–86, 2018.
[10] E. De Cristofaro, P. Gasti, and G. Tsudik, "Fast and private computation of cardinality of set intersection and union," in *International Conference on Cryptology and Network Security*. Springer, 2012, pp. 218–231.
[11] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, "Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing," *Science*, vol. 368, no. 6491, 2020.
[12] "Contact tracing: critical method to control the spread of covid-19," https://www.aarogyasetu.gov.in/wp-content/uploads/2020/11/mygov-999999999708639531.pdf, May 2020.
[13] "Thailandplus," https://thailandplus.in.th/en/, 2020.
[14] S. Wray, "South korea to step-up online coronavirus tracking," https://www.smartcitiesworld.net/news/news/south-korea-to-step-up-online-coronavirus-tracking-5109, Mar. 2020.
[15] M. Gad-Nowak, "Covid-19: Poland launches an official tracking app," https://www.natlawreview.com/article/COVID-19-poland-launches-official-tracking-app, Apr. 2020.

[16] Y. Luo, C. Zhang, Y. Zhang, C. Zuo, D. Xuan, Z. Lin, A. C. Champion, and N. Shroff, "Acoustic-turf: Acoustic-based privacy-preserving covid-19 contact tracing," *arXiv preprint arXiv:2006.13362*, 2020.

[17] R. L. Rivest, J. Callas, R. Canetti, K. Esvelt, D. K. Gillmor, Y. T. Kalai, A. Lysyanskaya, A. Norige, R. Raskar, A. Shamir *et al.*, "The pact protocol specification," *Private Automated Contact Tracing Team, MIT, Cambridge, MA, USA, Tech. Rep. 0.1*, 2020.

[18] J. Chan, D. Foster, S. Gollakota, E. Horvitz, J. Jaeger, S. Kakade, T. Kohno, J. Langford, J. Larson, P. Sharma *et al.*, "Pact: Privacy sensitive protocols and mechanisms for mobile contact tracing," *arXiv preprint arXiv:2004.03544*, 2020.

[19] C. Watch, "Covid watch sourcecode," https://github.com/covid19risk/, 2020.

[20] T. Coalition, "Tcn protocol for decentralized, privacy-preserving contact tracing," 2020.

[21] R. Raskar. (2020, Jun.) Covid-safepaths. https://github.com/Path-Check/covid-safe-paths.

[22] H. Xu, L. Zhang, O. Onireti, Y. Fang, W. J. Buchanan, and M. A. Imran, "Beeptrace: Blockchain-enabled privacy-preserving contact tracing for covid-19 pandemic and beyond," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3915–3929, 2020.

[23] W. Lv, S. Wu, C. Jiang, Y. Cui, X. Qiu, and Y. Zhang, "Towards large-scale and privacy-preserving contact tracing in covid-19 pandemic: A blockchain perspective," *IEEE Transactions on Network Science and Engineering*, 2020.

[24] K. C. Toth and A. Anderson-Priddy, "Self-sovereign digital identity: A paradigm shift for identity," *IEEE Security & Privacy*, vol. 17, no. 3, pp. 17–27, 2019.

[25] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC)*. IEEE, 2017, pp. 1–5.

[26] T. H.-J. Kim and J. Lampkins, "Ssp: Self-sovereign privacy for internet of things using blockchain and mpc," in *2019 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 2019, pp. 411–418.

[27] M. J. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection," in *International conference on the theory and applications of cryptographic techniques*. Springer, 2004, pp. 1–19.

[28] J. Sun, R. Zhang, and Y. Zhang, "Privacy-preserving spatiotemporal matching," in *2013 Proceedings IEEE INFOCOM*. IEEE, 2013, pp. 800–808.

[29] D. Boneh, "The decision diffie-hellman problem," in *International Algorithmic Number Theory Symposium*. Springer, 1998, pp. 48–63.

[30] D. Wang, C. Huang, X. Shen, and N. Xiong, "A general location-authentication based secure participant recruitment scheme for vehicular crowdsensing," *Computer Networks*, vol. 171, p. 107152, 2020.

[31] R. Zhang, R. Xue, and L. Liu, "Security and privacy on blockchain," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–34, 2019.

[32] Y. Zheng, M. Li, W. Lou, and Y. T. Hou, "Location based handshake and private proximity test with location tags," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 4, pp. 406–419, 2015.