Exploring New Paradigm for COVID-19 Contact Tracing Through the Eye of IoT: A Blockchain Approach

Amar A. Rasheed¹, Mingkui Wei², Rabi N. Mahapatra³

^{1,2}Department of Computer Science, Sam Houston State University, Huntsville, TX 77320 USA

³Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843 USA

Abstract— Existing contact tracing technologies for the novel Coronavirus were built to take advantage of the user's cellphone embedded features. Specifically, exposure notification system and Bluetooth proximity were widely adopted by many of today's Covid-19 contact tracing technologies. This paper proposes a new technological paradigm that supports contact tracing capability for the Covid-19 virus. The proposed system's models can be used to determine potential virus outbreaks within a controlled community. In this paradigm, we primarily focus on the enrichment of contact tracing capability for communities that have limited access to smartphones (e.g. children's schools, senior homes, and factories/warehouses). Several technological and crypto innovations were incorporated into the proposed system's model design: (i) Internet-of-Things contact tracing data logger (ii) Two-factor Authentication (iii) User's anonymity and data integrity of the contact tracing data (iv) detection and localization of outbreaks in a community. Contact tracing data for Covid-19 are collected using users-authorized IoT devices. Accesses to the Covid-19 contact tracing system were established by a new two-factor authentication. The proposed authentication scheme combined public key crypto, a blockchain validation approach, and a random token generator. Meanwhile, user anonymity was achieved via the exchange of randomly generated IDs, computed by light-weight crypto hash for supporting data logging of Covid-19 traces. To preserve the integrity of Covid-19 contacts tracing data, we have adapted a blockchain methodology. The proposed approach protects user data against data injection and data modification attacks. Finally, we have analyzed the resiliency of the proposed system's models against two well-known IoT-based attack models: (i) Sybil attack (ii) Replay attacks. Two probabilistic models were introduced in this research that present system resiliency under different scales of threats.

I. INTRODUCTION

Covid-19 has been declared as the fifth pandemic following the Spanish Flu pandemic that occurred during the year 1918. The first case of a novel Coronavirus

Fig. 1. Confirmed cases of Covid-19 for the period Dec. 2019 to Jul. 2020 [39]

(Covid-19) was reported in Wuhan city [1], [2], [3] China and it is traced back to December 1, 2019. In late December 2019, more people have been infected by the novel virus, most of the patients were experiencing mild to severe Upper Respiratory Infection-like symptoms [4], [5], [6]. Patients identified with Covid-19 appeared to have a high fever, cough, shortness of breath, new loss of test or smell, nausea or vomiting, and diarrhea [7]. It has been recorded that symptoms might appear between 2-14 days after a human being is exposed to the novel virus. Soon after the virus outbreak in china and by mid-January 2020, it became clear that the world is dealing with a new pandemic that is driven by a highly contagious disease. In March 2020, the World Health Organization (WHO) has declared the novel Coronavirus as a pandemic. Until a safe and effective vaccine is made available, people will remain vulnerable to getting infected with Covid-19.

For the past few months as we rolling through this pandemic (see figure 1), government agencies including commercial sectors in many countries including Asia and Europe has been heavily investing in the development of new protective measurements [8], [9], [10], [11], [12] that minimize human exposure to the virus, some of these countermeasures, including shutting down the economy to minimize social interactions, wearing facial covers [13] to prevent respiratory droplets that carry the virus from being landed in the mouth or noses or possibly inhaled into the lung of a healthy person [14-18]. Other prevention methods mainly focused on the development of new technology that supports contact tracing capabilities [19], [20], [21] and identifying if a person has been possibly exposed to the virus through the establishment of close contact with an infected person. The new technology has been deployed as a contact-tracing app for mobile platforms and cellphones [22], [23], [24], [25], [26]. Many of the Covid-19 tracing apps rely on geotagging and Bluetooth technology, such apps run on the user's cellphone, generally in the background, and freely exchange randomly generated IDs with other cellphones that the user made close contact. For Example, Care 19 [27] is a contact-tracing app developed to provide users with the capability of accessing information related to whether a user has been near a positive Covid-19 subject. The app uses Bluetooth proximity technology developed by Apple and Google exposure notification systems. The system keeps a track of anonymous keying information that a user encountered during the day. Keying information is randomly computed and securely exchange with other systems that the user contacted. To ensure user privacy, random IDs are recomputed every 10-20 minutes. Although Care 19 contact tracing app provides a method for determining if a user has been exposed to the Covid-19 virus or not, the app can be used by a malicious actor to track user location within a 20-minute time frame. As Care 19 app support no access to geotagging data, adversaries can still utilize the above technology to tag a user location using relative distance information measured between the victim device and the adversary via Bluetooth proximity sensing, combined with the adversary GPS data, an adversary will be able to collect user approximate locations data for 10-20 minutes before the victim's device reset its random ID (see figure 2).

Fig. 2. Tagging user approximate location using Bluetooth proximity sensing data

Another shortcoming of the above contact-tracing technology is that the Care 19 will be incapable of preserving the integrity of the contacts tracing data. Contact tracing data or random IDs will be stored onboard the mobile device. The mobile device maintains a periodic check with the Care 19 server to identify if at least one of its collected random IDs is associated with a positive Covid-19 case or not. Since contact tracing data are stored on the device memory, an adversary with remote access to the phone will be able to inject malicious contact tracing data modifying the integrity of the trace, which hence makes Care 19 an unreliable system. To the best of our knowledge, Care 19 provides no mechanism that can protect the user from such an attack. In this paper, we will introduce an innovative Covid-19 tracing capability that is cost-effective and does not require the user to have access to a smartphone. The proposed technology supports contact tracing capability on the community level rather than the general population. Although the proposed technique has been developed to target tracing of positive Covid-19 cases within a community, it can be easily adapted for public use. In this research, our focus is to allow users of a community to be alerted when they made contact with a positive Covid-19 subject within their community. In this paradigm of community, a community is defined by factory workers, students, staff and faculties of academic institutions, employees of large government organizations, and commercial corporations. The proposed technology will support the following capabilities:

- *Internet of Things (IoT)-Covid-19 Data Logger Platform.* An IoT-based platform that utilizes Bluetooth proximity sensing capabilities to detect and log contacts tracing data, in the case of a user who has been in close contact with another user within the same community. The proposed platform will act as a replacement for the smartphone with a Covid-19 contact tracing app installed on it. IoT-based platforms will be controlled and managed by the community's organization. Devices will be randomly assigned to the community's users.
- *Two-factor Authentication Mechanism.* Streaming contacts tracing data logs and accessing the Covid-19 alert system is controlled via two access control mechanisms. The first access control method is established via a Public key cryptosystem [28-30] combined with a blockchain validation approach. With this method, both the IoT-Covid-19 data logger device and the Covid-19 contact's tracing system will be able to authenticate each other securely. The validation of user to IoT-Covid-19 data logger device association is maintained via the deployment of a computer/mobile platform app that is capable of creating a new random token each time a user tries to access the Covid-19 alert system. With each token randomly generated at the user's app, the same random token is computed at the Covid-19 alert system. Each user's app is implemented with a random function generator that is unique to each user in the community and is directly linked to the user's account.
- *User Anonymity.* To minimize the exposure of user's random IDs over the system and to prevent malicious users within the same community from utilizing their IoT-Covid-19 data logger devices to track other users' locations. We achieved user anonymity via the implementation of a secure one-way cryptographic hash engine onboard the IoT-Covid-19 data logger device. Hash keys that are preloaded into each IoT device memory are available to the Covid-19 alert system through the user account data. This technique will enable obfuscation of the IoT device's real identification information each time the device made contact with another nearby device. For each user's device, a new random device Id is being generated during each contact tracing data logging event. To exploit the above system, malicious users need to solve a moving target problem.
- *Covid-19 Contact Tracing Data Integrity:* To preserve the integrity of each user's contacts tracing data, we have adopted a block chaining methodology to protect user data against data injection and data modification attacks.

A new block of contacts tracing data can be added to the user's blockchain only if the user's IoT device is capable of validating the computation of a previous block's hash value.

- *Covid-19 Outbreak Localization within the Community.* In the proposed framework, within a community, we consider the monitoring of places with high human traffic. For example, in the college campus community, places that could potentially cause a virus outbreak to include college dormitories, recreation and athlete center, lecture rooms, and dining hall. Community areas that have a high volume of human traffic are equipped with IoT-based Covid-19 area data logger devices. Such devices provide daily reports for community users' traffic volumes. Such information alerts the users regarding areas that could lead to a potential virus outbreak in the community. Community users could utilize this information to avoid virus hotspot locations.

This paper is organized as follows. Section II provides preliminaries and related backgrounds used in some subsequent sections of the paper. Section III introduces the proposed design of the IoT Covid-19 data logger platform. In Section IV, we present a two-factor authentication method that's based on a random token's app generator combined with the public key cryptosystem. Section V proposes a new approach that supports user anonymity. A novel approach that is based on blockchain technology is presented in section VI. The new approach will be used to preserve the integrity of each user's contact tracing data. Section VII describes a methodology for the localization of outbreaks within a community. Three threat models will be presented in section VIII. In section IX, we analyzed the resiliency of the proposed scheme against two threat models. Section X concludes the paper and presents future research directions.

II. PRELIMINARIES

The proposed approaches herein have been built based on the utilization of two crypto technologies: (i) blockchain, and secure hash.

A. Blockchain

Blockchain technology [31], [32], [33] is a decentralized system of shared and immutable ledgers. The system is capable of recording transaction records. Transaction records or sometimes called blocks are timed stamped, secured, and bonded to other blocks in the chain. Each block in the chain is cryptographically linked to the previous block in the chain block.

Blockchain has no central authority, making the system immutable against block's injection and block's modification attacks. There are two types of blockchain networks: (i) public blockchain networks (ii) private blockchain networks [34]. In a public blockchain, anyone can participate in the system, and add a new block to the chain. Transactional data are often can be accessed by anyone on the system. Such networks are widely deployed for Bitcoin, they have weak security measurements, and with little to no privacy. On the other hand, private blockchain networks are a decentralized peer-to-peer system that is usually managed and controlled by a single organization. The system is built based on peer-to-peer network communications, only users of the network can participate in a private chain. Transactional data records are kept secure and private with no public access. Participants of the network run a consensus scheme and able to access the network's shared ledgers. The latter brings more trust and confidence with an enhanced tracing capability compared to a public blockchain network.

Since our proposed approach support Covid-19 contacts tracing capability on the community level. We have considered the adaption of the private blockchain network model during this work. In our approach, we have assumed that Covid-19 contacts tracing data of the user's community are recorded via a single private blockchain network. Multiples private ledgers will be utilized by the community organization to maintained and managed hundreds to thousands of participants' contact tracing transactional blocks. Covid-19 contact tracing blocks are linked together to form an individual blockchain for each user in the community. These individual blockchains provide tracking information for 14 days which is the incubation period of the novel Coronavirus. The integrity of the user's contact tracing data can be achieved via the deployment of a single private blockchain network with multiple ledgers.

B. Secure hashing

In this effort, we have considered the use of cryptographic hash to achieve the following capabilities:

1) *Anonymity*: To eliminate the exposure of a user ID during the contact tracing process, new users' IDs are randomly generated onboard their IoT Covid-19 datalogger devices for each contact tracing event. We rely on the deployment of one-way cryptographic hash [35-38], such as SHA-1. We proposed the implementation of an asymmetric cryptographic hash engine on each Covid-19 datalogger device deployed in the community, including the Covid-19 alert system.

In our approach, a user's id is randomly computed at the device level, where each user id is being obfuscated via the addition of random token α and applying secure hashing on the obfuscated id $H(id \parallel \alpha)$. The hash value will serve as the device's new id, keeping the device's real id private. During a contact tracing event, each user device will share it's computed hash values $H_j (id_i \parallel \alpha_i)$ along with randomly generated token values α_i over the network. The Covid-19 alert system parses the computed hash values and tokens to compute users' devices' real ids and therefore, the system will be able to recognize each user blockchain for logging of new blocks.

1) *Hash-based Authentication*: We propose a method that enables secure bonding between each user's community, and his/her assigned IoT-Covid-19 data logger platform. To avoid the injection of malicious blocks into the network, the authenticity of each assigned IoT Covid-19 data logger platform within the community is verified before logging Covid-19 tracing data blocks. Verification of IoT platforms is implemented via the deployment of user-specific token-computing apps that are generated and preconfigured during the users' accounts creation step. A user-specific app is preconfigured with the user's device identification. It supports symmetric key encryption for secure communication between the app and the user IoT Covid-19 data logger device. In our design, a secure user-specific app acts as a relay to transfer contacts tracing data between the user device and the Covid-19 alert system. The user-specific token-computing app provides another form of verifying the authenticity of the data logger platform. The app is implemented with a crypto hash SHA-3 engine, the same hash engine is also deployed on the Covid-19 contact tracing server. Hash values or app tokens computed at the users' apps and the server platforms will be compared for users' devices verifications. With this approach, Our Covid-19 contact tracing system will be capable of verifying the user and its IoT data logger platform cryptographically.

Fig 3. IoT Covid-19 data logger devices assignment and deployment in a community

III. INTERNET OF THINGS (IOT)-COVID-19 DATA LOGGER PLATFORM

Existing Covid-19 contacts tracing technologies have been fully relying on the user's ability for having access to a smart mobile platform. These smart devices support data logging and provide detection of events that are triggered due to physical contacts made by the user. In communities where users do not have access to smartphone technology (e.g. community like senior homes, and schools), such users' might not be able to utilize the above technology for tracing positive Covid-19 cases within their communities.

Fig. 4 Apps deployment a device's assignments

Therefore, we proposed the development of new cost-effective, power-efficient, and light-weight Covid-19 contact detection and data logger IoT-based platform, the new technology provides continuous monitoring of users in a community to determine whether a user has been in physical proximity with another user in the community or not. The proposed IoT platform utilizes its embedded

Bluetooth technology and proximity sensors to detect a contact tracing event that occurred between users in the community. The onboard Bluetooth provides a secure exchange of randomly generated users' ids. We use the IoT embedded proximity sensors module to provide an estimation for distance measurement between two users which are detected by the IoT contacts tracing platform within a 6feet distance.

In this research, we are presenting a system that can provide anonymous contact tracing data for the novel Covid-19 virus within a community. Small communities, like schools, senior care centers could utilize the proposed

Covid-19 contacts tracing system with hundreds of IoT data logger devices are being deployed (see figure 3). Meanwhile, communities with a large population, like college campuses or factory workers, where we consider a system with thousands of IoT Covid-19 data logger platforms are being assigned to these user's communities.

A. IoT platforms assignments: Each IoT Covid-19 data logger platform, is preconfigured with a unique identification number. Identification values are hardware coded into each IoT device to prevent hardware cloning. During the devices' assignment steps, each user in the community will be assigned to a single IoT Covid-19 data logger platform. Hardware-based identification values for the assigned IoT platform will be used to create users' accounts in the system. To prevent malicious users' from being able to tag and identify other users, Hardware-based ids (users accounts ids) are obfuscated by a cryptographic hash engine onboard the IoT device before being shared with other IoT devices. Randomly generated ids are computed on the fly by each IoT platform during the contacts tracing process and communicated over the system, keeping users' accounts ids secret.

B. Distribution of user-specific token computing app: During the user subscription process, each user will be able to download a token-computing app (see figure 4), each downloaded app is uniquely preconfigured and coded with the user's account id and asymmetric hash key shared with Covid-19 servers.

Fig 5: Covid-19 contact tracing system architecture

During the logging of contacts tracing data into the Covid-19 alert system, randomly generated tokens are computed by the users' apps and then entered by the users into Covid-19 contacts tracing systems to provide secure access and initiate the data logging processes. With the proposed technique, users' apps are used to verify the authenticity of their IoT data logger platforms. Meanwhile, computed tokens with user account sign-in information are used to verify both the user and his/her assigned IoT data logger device.

IV. THE PROPOSED AUTHENTICATION SCHEME

In this paper, we proposed a two-factor authentication technique that supports historical traces for every Covid-19 data logging event. We combined public key crypto and validation of a previous block within the IoT device blockchain for authentication. We have considered the deployment of a lightweight Elliptic Curve Crypto (ECC) engine onboard the IoT device and implemented it on the user-specific token computing app. Each time, a user requires to access his/her Covid-19 contacts tracing data, both the user's IoT device and its token computing app employs the ECC engine during the first stage of the authentication process. During the second stage of the authentication method, the user's app computes a new token using the blockchain approach, the new token is being presented as a new block in the token's blockchain.

At the user's app, a data record containing the new token's value, and the new token's time of creation is computed and sent to Covid-19 servers. This new data records along with the value of the new token's block inputted by the user during sign-in will be used by Covid-19 contacts tracing servers to validate the content of a previous block in the token's blockchain. In the proposed scheme, we have assumed that the IoT data logger device can establish a secure channel with the community's server, only via the user-specific token computing app. The token computing app acts as an extra layer of protection against Denial of Service Attacks that could be launch by an adversary with a large number of compromised IoT data logger devices. Also, the token computing app provides immunity against attackers that utilize the user's stolen IoT data logger platforms to inject malicious contacts tracing data into the system. In our setting, we have assumed that each user's IoT data logger device is equipped with a Bluetooth communication engine and with no internet capability. The Bluetooth communication engine is mainly used for (i) triggering contacts tracing events of Covid-19 (ii) establishing a secure communication channel between the IoT platform and the user-specific token computing app.

The proposed technique can provide a full trace for all Covid-19 successful authentication and data logging transactions that are generated per IoT device.

Each device's authentication and data logging events are recorded in the form of a blockchain. We have adopted a blockchain approach to preserve the integrity of each IoT device authentication and data logging attempt. Since each user's account in the community has been linked to a unique token computing app and a single pre-configured IoT device, unique records of authentication blocks are computed and added to the device's blockchain.

A. Authentication process

During the authentication step, the following interactions will take place between a user-specific token computing app and the Covid-19 contacts tracing system's server (see figure 5).

- 1) A user's IoT Covid-19 data logger device sends a challenge-response authentication message $m_i = E_{Pub_i}(noun)$ to the user's token computing app via a Bluetooth connection established with the app's host machine.
- 2) A user's specific token computing app receives the challenge m_i and computes $c_i = D_{Pr_i}(m_i)$. The app then sends the computed challenge c_i back to the IoT Covid-19 data logger device via Bluetooth.
- 3) The user's IoT Covid-19 data logger device will compare if the received message c_i is equal to the challenge $noun$. Using the public key infrastructure, if $c_i = noun$, both the user's token computing app and the IoT's data logger device will be able to authenticate each other successfully.
- 4) During this step, the token computing app computes a new token block $B_{T(i)} = hash(B_{T(i-1)} || t_i || Id_x || TK_i)$, where $B_{T(i-1)}$ is the previous block in the user's token blockchain, t_i is the creation time of block $B_{T(i)}$, TK_i is a randomly generated token, and Id_x is the hardware coded id of the user's IoT platform.
- 5) A Covid-19 contacts tracing server and the user's token computing app will use a PKI, similar to step 1 of the authentication process. The latter provides secure authentication and establishes a secret channel for transferring contacts tracing data between a Covid-19 contacts tracing server and the user's token computing app.
- 6) The user's token computing app sends the ciphertext messages $E_{Pub_i}(TK_i)$, and $E_{Pub_i}(t_i)$ to the server.

7) Device-to-App authentication was provided via step 1 and step 2 of the protocol. App-to-Server authentication was established via step 5. In this step, Device-to-Server authentication was maintained through the integration of an engine that allows a user to input his/her randomly generated token block $B_{T(i)}$, plus his/her security credentials to access the community's Covid-19 contacts tracing servers.

8) Finally, a server will try to validate the integrity of the new token block $B_{T(i)}$ by recomputing it based on the stored value of the previous token's block $B_{T(i-1)}$, the timestamp data t_i , the device's hardware-coded Id_x , and the current token's value TK_i .

V. USER ANONYMITY

Our proposed methodology for tracing Covid-19 positive cases in a community fully relies on the free exchange of users' identifications with other users in the community. As described earlier, each user in the community has been equipped with an IoT platform and the assigned IoT platform is used by the user to log and provide contacts tracing information daily.

Hardware-coded ids embedded in the IoT platforms are used to identify different users in the community. To trace users' contacts in a community, users are required to surrender their ids to other users in the community. This overall system behavior of exchanging users' ids over the system could be exploited by a malicious user to build a daily digital trail of the user activities, like with whom he/she has been in close contact. Also, an attacker could utilize GPS technology to acquire a daily trace for location estimation of the victim. Another possible attack will be spoofing users' ids to launch highly sophisticated attacks like DDOS to inject malicious contact tracing data into the system. Therefore, we have considered a new approach that obscures the user's hardware-coded ids during the ids exchange process. As a user become near another user in the community, obfuscated users' ids are computed and exchanged securely between the users' IoT platforms. During each contact tracing event, the proposed system ensures that no user's id is being used twice. Each contact tracing event will be presented with a freshly generated unique id, making it impossible to trace a user activity based on gathering his/her contact tracing data logs.

A. Generation of ids

The following describes our proposed approach for obscuring users' hardware-coded ids before exchanging them with other IoT devices. When two IoT Covid-19 data loggers contact each another via Bluetooth proximity exposure capability, each IoT device computes a randomized id based on its assigned hardware-coded identification information. A randomly generated ID_x is computed at IoT device x using a cryptographic hash, where $ID_x = H_x(Id_x || \alpha_x)$, Id_x is device x hardware-coded id, and α_x is a randomly computed noun to garble the device's true id. For example, two IoT platforms x and y that are in the process of tagging each other, the two devices will start computing their random ids $ID_x = H_x(Id_x || \alpha_x)$, and $ID_y = H_y(Id_y || \alpha_y)$. Then IoT device x will exchange its newly computed id ID_x , plus the randomly generated α_x with IoT device y . Similarly, IoT device y will exchange its computed random id ID_y and it computed noun α_y with device x . Since a randomly generated noun α_x is used to garble each device's hardware-coded id before transmission, malicious users will not be able to tag a nearby IoT device id. To successfully log contact-tracing data into the community's Covid-19 servers, servers must be capable of extracting the true users' identification information from the stream of data that is collected by each IoT platform. For a Covid-19 contact tracing block, tagged with an obscured id ID_j , a community's Covid-19 server needs to compute multiple hash values $H_0(Id_0 || \alpha_j), H_1(Id_1 || \alpha_j), \dots, H_j(Id_j || \alpha_j)$ until it finds a match where $H_j(Id_j || \alpha_j) = ID_j$. Then the server will be able to extract the hardware-coded id Id_j that is garbled during a contact tracing event. In the proposed approach, we assumed a system with multiple Covid-19 servers that are capable of running hundreds of user's id extraction engines simultaneously. As users start transferring their daily contacts tracing logs for Covid-19, the system securely extracts users' ids from Covid-19 data logs, where extracted users' ids are cross-referenced with a database system to determine whether a user has been near a Covid-19 positive case or not.

VI. PRESERVING DATA INTEGRITY FOR COVID-19 CONTACT TRACING BLOCKS

We achieved data integrity for Covid-19 contacts tracing records via the employment of a blockchain methodology (see figure 5). Users make daily movements within their community, they might come in close contact with other users.

The physical distance between any two users in the community is being continuously monitored by the users' IoT devices. Users' devices use the Bluetooth proximity exposure capability to provide an estimation of the distance between two users in the community. A comparison of the calculated distance estimations with a preconfigured threshold value (6 feet) are handled onboard the IoT platforms on the fly. If the calculated distance estimation between users is less than 6 feet, the two users' IoT platforms will initiate the process of computing Covid-19 contact tracing data. To avoid the injection of malicious data blocks into a user's Covid-19 contact tracings blockchain, we have considered an approach that utilizes the user-specific token computing app to store the user's contact tracing's previous block data. Contact tracing data blocks will be computed by the user-specific token computing app. Data of previous blocks will not be stored on the IoT devices; this will prevent a malicious user from being able to compute a new block. The following describes the steps (see figure 5) that two IoT devices, x , and y with hardware-coded Id_x and Id_y needed to compute their contact tracing data:

- IoT devices x and y will first compute their randomized ids as $ID_x = H_x(Id_x || \alpha_x)$, and $ID_y = H_y(Id_y || \alpha_y)$ respectively, where Id_x , Id_y are hardware-coded ids of IoT devices x and y . The values α_x and α_y are the randomly computed noun at device x and y respectively.
- IoT devices x and y will then exchange their computed ids with each another, plus their randomly generated nouns α_x and α_y . All transactions are timestamps.
- Records of contact tracing are computed and stored into each user's IoT Covid-19 data logger as users move within their community.
- As the user tries to transfer their collected contact tracing records. Users initiate the two-factor authentication process that will provide Device-to-Server validation. After a successful authentication is established between the server and the user IoT platform, the IoT device will start relying on its recorded contacts tracings data to the user-specific token computing app.
- The user-specific token computing app uses the last contact tracings block. Such a block is computed during the last transfer and it is employed as a previous block during the computation of successive blocks.

As new contacts tracings data is being transferred from the IoT device to the user's token computing app, new contacts tracings blocks are computed as $B_{(i)} = \text{hash}(B_{(i-1)} || t_i || ID_x || TK_j)$, where B_i is the new contact tracing block, and B_{i-1} is the previous block in the chain. The new contact tracing data block is sent to the Covid-19 server along with the block's contacts tracings data, t_i , ID_x , and TK_j , where TK_j is the randomly computed token during transfer j .

VII. LOCALIZATION OF COVID-19 OUTBREAKS IN A COMMUNITY

In this section, we introduce a framework for localizing possible outbreaks of the coronavirus within a community. As illustrated in figure 6, we considered the deployment of the user's traffic monitoring system in areas that are exposed to a large scale of human traffic. The proposed system will be capable of providing real-time feedback information to users in case an outbreak has been detected in one of the areas that are under monitoring. For example, on a college campus, classroom building, including hallways areas can be equipped with Covid-19 outbreak localization systems. Such systems can be deployed on buildings entrances. Each area is controlled by a network of wireless-enabled IoT sensing modules and multiple edge computing systems. The area's sensing modules are programmed to establish secured channels with any nearby authorized data logger device. After establishing a secure channel with a nearby sensing unit, a user's data logger device will be able to securely transfer its user hardware-coded id to the sensing module. Meanwhile, hardware-coded ids captured via the area's sensing modules will be communicated directly to the edge computing system. Periodically, edge computing systems are being uploaded with real-time data of all Covid-19 positive cases that are identified within a community. Such data can be used to detect the presence of an outbreak and send warning messages over the Covid-19 contact tracing network systems, alerting users to avoid areas with virus outbreaks. Algorithms for Covid-19 contact tracing and detection of virus outbreak will run continuously on each area's edge computing systems. As new users entered an area, their captured hardware-coded ids will be computed by the edge computing systems and cross-referenced against a database of Covid-19 positive cases. If more than one user is identified by the edge computing system as being coronavirus positive, a warning message will be transmitted over the entire community, announcing the detection of a possible virus outbreak.

In our proposed Covid-19 outbreaks localization system, we have assumed that the area's sensing modules are integrated with a dual wireless communication system. Each sensing module is equipped with a Bluetooth communication chip and a Wi-Fi chip. Area's sensing modules and the users' data logger platforms will utilize their Bluetooth embedded capabilities to (i) establish a secure channel (ii) enable secure transfer of the user's hardware-coded ids into the area's edge computing systems for processing. In this framework, the area's sensing modules act as authenticator nodes for user's data logging devices.

Fig. 6 Detection system for localizing virus's outbreaks in a community

In addition to acting as authenticator nodes, they support secure relying of the user hardware-coded ids to the edge computing systems. Sensing modules and the area's edge computing system will communicate securely over a Wi-Fi channel. Therefore, users' data loggers will be fully isolated from the public network, protecting them from malicious actors that target the integrity of the collected contact tracing data.

VIII. THREATS MODELING

In this paper, we introduce three threat models that enable an attacker to inject false contact tracing data into a user's Covid-19 data logger device. We have classified such threat models as data modification and injection attacks that are mainly targeting the reliability of the proposed Covid-19 contact tracing system. To study the resiliency of the above system, we have considered the following threat models:

Sybil attacks: A Sybil attack is accomplished via the cloning of multiple users' Covid-19 data logger devices (see figure 7). In our proposed system, users are assigned to IoT platforms with unique hardware coded id for logging contact tracing data. We have considered an attacker that is capable of making multiple clones from compromised users' data loggers. Cloned IoT data loggers will be randomly deployed on the community by the attacker. A user's with Covid-19 data logger devices that are within the communication range of a Sybil node will exchange its garbled ids with the attacker device. False contact tracing data will then be injected into the user's account.

Fig. 7 Sybil attack with a single spoofed id

Replay attacks: In this type of attack, an adversary will have multiple malicious nodes deployed over the community. Such a threat can be implemented by an insider. As shown in figure 8, a user with an authorized Covid-19 data logger can utilize his device to inject faulty data into other users' Covid-19 data logger devices. The attack takes place when a malicious user alters his/her device's embedded software implementation. Instead of sending a garbled value of the device's true hardware coded id, the infected device searches its internal memory looking for a garbled id that is collected during previous contact tracing events. The malicious actor uses a spoofed id when making close contact with other users in the system. This will maliciously modify the Covid-19 contact tracing data for the victims. The injection of false contact tracing data makes the Covid-19 alert system unreliable. Although a user has not been near an infected individual, a user under a replay attack could be alerted due to a system's miss-detection.

Stolen Covid-19 data loggers: We have considered the possibility that a lost or a stolen Covid-19 data logger can be utilized by an attacker to inject malicious contact tracing data into a user account.

In these types of attacks, an attacker uses a stolen/lost data logger device to establish close contact with other Covid-19 data logger devices in the community. Modifying the integrity of a user Covid-19 contact tracing logs. An inaccurate trail of contact tracing data will be computed by the attacker for the user with the lost/stolen device.

Fig. 8 Mobile replay attack with multiple spoofed ids

IX. SYSTEM RESILIENCY AGAINST DATA MODIFICATION AND DATA INJECTION THREATS

In this research, we are mainly interested in providing a methodological framework for the development of a reliable Covid-19 contact tracing system, where users are capable of accessing accurate tracing information related to Covid-19 within a community. Therefore, to study the reliability and the visibility of the proposed Covid-19 contact tracing system, we have considered three possible types of attacks (i) Sybil attacks (ii) Replay attack (iii) stolen/lost Covid-19 data logger device. Such attacks could be mapped directly to the data modification/data injection attack class. At a given time interval, a community exhibit random occupancy by its users. Users with assigned Covid-19 data loggers randomly traverse the community. In this paper, we have considered the modeling of community areas that show a high volume of human traffics. For example, warehouses, college dormitories, and dining halls. Such an area usually has a steady human density for a given time interval. We investigated how Sybil/replay threats target the localizations of users' ids within a community's area.

To model the resiliency of the proposed Covid-19 contact tracing system, we assumed a community's area with user's density σ , and a Sybil node's or replay attacker's node with a coverage area of πr^2 , where r is the attacker's node communication range. The estimated fraction of users' data loggers that are within the Sybil/replay attacker's node area is $\beta = \sigma \pi r^2$.

A Sybil attack is being modeled as a Poisson Distribution, where the probability of having k users' data loggers under attack can be estimated as:

$$p(k) = \beta^k \frac{e^{-\beta}}{k!} \cdot q(k),$$

where $q(k) = q^k$ represents the probability of having k data loggers injected with a spoofed id during a Sybil attack. A spoofed id used by the Sybil device will be successfully added into a user's contact tracing data only if the id does not belong to a data logger device that is currently within the victim communication range. The probability $q(1)$ of having a spoofed Id_r being successfully injected into a user data logger can be measured as:

$$q(1) = (\omega - \delta) / \omega$$

Where ω represents the total number of data logger devices deployed in the community, and $\delta = \sigma\pi R^2$ is the total number of data loggers that are currently within the victim's communication range $2R$.

In the case of a replay attack, we have estimated the probability of having k Covid-19 data loggers being successfully injected with n spoofed ids, where the injected ids are held by devices outside the communication ranges of victims' devices. Covid-19 contact tracing data for each of the victims' devices will be poisoned with malicious data blocks, making Covid-19 tracing capability unreliable for such users. We have measured the probability of a replay attack as:

$$z(k) = \beta^k \frac{e^{-\beta}}{k!} \cdot q(k)^n$$

We have analyzed the probability of having k Covid-19 data logger devices under a Sybil attack. In our threat modeling for the Sybil attack, we have assumed a single spoofed user's id is used by the attacker to launch Sybil's node attack. We have estimated the probability $p(k)$ under various community sizes {250, 500, 750, 1000, 1500} users. Also, we have determined the probability $p(k)$ under various attacker's communication range, r {10m, 15m, 20m, 25m}. Figure 9 shows the probability of having k Covid-19 contact tracing loggers being injected with a single spoofed id. For a given community size, as we increase k , the probability of a successful Sybil attack decrease. Meanwhile, as we increase the size of the community from 250 users to 1500 users as shown in figure 9(a), the probability of having 9 data loggers under a Sybil threat increases by 36.27, and this due to an increase in devices density within the community.

The average number of data loggers being under Sybil threat increases as we elevate the number of devices deployed in a community. As illustrated in figure 9(a), for a given probability of attack $p(k) = 0.11825$, the number of devices under Sybil threat increases from 4 to 8 devices for community sizes of 1000 users and 1500 users respectively. For a 0.05 probability of attack and a community size of 1000 users, figures 9(b), 9(d) show an increase in the number of devices under attack from 10 to 39 devices for changes in the attacker communication range from $r = 15m$ to 25m.

Figure 10, presents the probabilities of k data logger devices being under a replay attack. We analyzed the probability $z(k)$ with various community sizes of {250, 500, 750, 1000, 1500} users and different numbers of spoofed ids, $n = \{5, 10, 15, 25\}$. Replay attack utilizes multiple spoofed ids to alter the contact tracing data for a user. Since multiple spoofed ids are used during the attack, the probability that a device within the attacker communication range is being injected with a malicious id became higher compared to a device being under the Sybil attack. Figure 10 (a) shows that with a probability, $z(k) \approx 0.1$, a replay attack size increases from 4 devices in a community with 1000 users to 9 devices for a community with 1500 users. Similar to the Sybil attack, as we increase the number of users per community from 1000 users to 1500 users, we have noticed there will be a 66% increase in the probability of 9 devices being under attack (see figure 10(a)). Finally, as illustrated in figures 10(a), 10(b), 10(c), and 10(d), the probability of a replay attack decreases as we increase the number of spoofed ids. For a community size of 1000 users, probability $z(k) = 0.098$ for $n = 5$, $z(k) = 0.067$ for $n = 10$, $z(k) = 0.058$ for $n = 15$, and $z(k) = 0.035$ for $n = 25$.

Our proposed contact tracing system for Covid-19 in a community provides full immunity against the lost/stolen. logger devices. Each user account on the Covid-19 alert system is protected by a two-factor authentication method. A public key infrastructure authentication mechanism is incorporated with a random token generator. To log contact tracing data into Covid-19 contact tracing servers, each user is required to supply his/her security primitives with the user's computed random token. The relaying of contact tracing data is performed via each user's token computing app. Moreover, injection of Covid-19 contact tracing logs from a stolen or a lost device required that an attacker has access to the user's assigned token computing application.

Fig. 9. The probability of k Covid-19 data loggers under Sybil attack with various community sizes

Fig. 10 The probability of k Covid-19 data loggers under replay attack using n spoofed ids and with various community sizes

Therefore, for an attacker with a stolen Covid-19 data logger to be able to inject faulty contact tracing data into a user's account with a stolen device, the attacker needs to hijack the user's account and be able to predict the next computed token value.

X. CONCLUSION

This paper proposed a novel contact tracing system for Covid-19, it supports anonymous contact tracing capability on the community level. Our system fully relies on the deployment of IoT platforms over the community's users. Users of a community will be capable of accessing their contact tracing information without the support of smartphone technology. Our system was specifically designed to capture users' contact traces during the day. It also provides a detection mechanism for localizing potential virus outbreaks within the monitored community. Various security schemes were incorporated into the design of the proposed system. In our system, users' IoT devices are authenticated by the Covid-19 contact tracing server via a two-factor authentication technique. The proposed authentication scheme combined public key crypto and validation of a previous block within the IoT device blockchain for authentication.

We have considered the deployment of a lightweight Elliptic Curve Crypto (ECC) engine onboard the IoT device and implemented a user-specific token computing app. User anonymity and data integrity of the user's contact tracing data were also integrated into the proposed system. We have achieved user's anonymity through the obfuscation of users' ids, using a random token and hash computing scheme, users' ids were garbled during the contact tracing exchanges process. For data integrity, we have considered the adaption of the private blockchain network model during this work. The integrity of the user's contact tracing data can be achieved via the deployment of a single private blockchain network with multiple ledgers.

We have considered three attack models in this work (i) Sybil attack (ii) replay attack (iii) lost/stolen IoT data logger device. Two probabilistic models were presented herein, the proposed models provide a probability estimation for having k Covid-19 data loggers being successfully injected with n spoofed ids, where the injected ids are held by devices outside the communication ranges of victims' devices. Also, we have analyzed the probability of having k Covid-19 data logger devices under a Sybil attack and with various community sizes.

REFERENCES

- [1] Ghinai I, McPherson TD, Hunter JC, Kirking HL, Christiansen D, Joshi K, et al. First Known person-to-person transmission of severe acute respiratory syndrome coronavirus2 (SARS-CoV-2) in the USA. *Lancet*. 2020.
- [2] [Online] Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) Feb. 16-24, 2020. Geneva. Available at <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>
- [3] [Online] Yanping Zhang. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19)-China 2020. *CDCWeekly*. Feb. 17, 2020, available at <http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51>
- [4] Liu J, Liao X, Qian S, Yuan J, Wang F, Liu Y, et al. Community Transmission of Severe Acute Respiratory Syndrome Coronavirus 2, Shenzhen, China 2020. *Emg Infect Dis*. 2020.
- [5] Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J, et al. A familial cluster of pneumonia associated with 2019 novel coronavirus indication person-to-person transmission; a study of family cluster. *Lancet*. January 23, 2020, 395(10223):514-523
- [6] [Online] Infection Prevention and Control of Epidemic-and Pandemic-prone Acute Respiratory Infections in Health Care. Geneva: World Health Organization. 2014. Available at https://apps.who.int/iris/bitstream/handle/10665/112656/9789241507134_eng.pdf;jsessionid=41AA684FB64571CE8D8A453C4F2B2096?sequence=1
- [7] [Online] COVID-19 Symptoms. Available at the CDC website at <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>
- [8] [Online] Considerations for public health and social measures in the workspace in the context of COVID-19: annex to considerations in adjusting public health and social measures in the context of COVID-19, May 10, 2020. Geneva: WHO. Available at <https://www.who.int/publications/i/item/considerations-for-public-health-and-social-measures-in-the-workplace-in-the-context-of-covid-19>
- [9] [Online] Key planning recommendations for mass gatherings in the context of the current COVID-19 outbreak: interim guidance, May 29, 2020. Geneva: WHO. Available at <https://www.who.int/publications/i/item/10665-332235>
- [10] [Online] United States Centers for Disease Control and Prevention. Interim infection prevention and control recommendations for patients with suspected or confirmed coronavirus disease 2019 (COVID-19) in a healthcare setting. *Coronavirus Disease 2019 (COVID-19)*, July 15, 2020. Available at <https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control-recommendations.html>
- [11] Rational use of personal protective equipment for coronavirus disease (COVID-19): interim guidance, April 6, 2020. Geneva: WHO, 2020. Available at [https://www.who.int/publications/i/item/rational-use-of-personal-protective-equipment-for-coronavirus-disease-\(covid-19\)-and-considerations-during-severe-shortages](https://www.who.int/publications/i/item/rational-use-of-personal-protective-equipment-for-coronavirus-disease-(covid-19)-and-considerations-during-severe-shortages)
- [12] [Online] Infection prevention and control during health care when novel coronavirus disease (COVID-19) is suspected or confirmed available at WHO: <https://www.who.int/publications/i/item/WHO-2019-nCoV-IPC-2020.4>, Jun 29, 2020.
- [13] Shuo Feng, Chen Shen, Nan Xia, Wei Song, Mengzhen Fan, Benjamin J Cowling, Rational use of face masks in the COVID-19 pandemic. *Lancet Respir Med*. May 2020. 434-436
- [14] Kai-Wang To, K. Tak-Yin Tsang, O., Chik-Yan Yip, C., Chan, KH., Wu, TC., Man-Chun Chan, J. Yuen, KY. Consistent detection of 2019 novel coronavirus in saliva, *Clinical Infectious Diseases*, Feb. 12, 2020.
- [15] Mittal R, Ni R, Seo J-H. The flow physics of COVID-19. *Journal of Fluid Mechanics*, Volume 894, July 10, 2020
- [16] Van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G. Gamble, A. Williamson, B. N. Tamin, A., Harcourt, J.L., Thornburg, N.J., Gerber, S.I., et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl Journal of Med*. March 17, 2020
- [17] Bourouiba L. Turbulent Gas Clouds and Respiratory Pathogen Emissions: Potential Implications for Reducing Transmission of COVID-19. *JAMA*. March 26, 2020
- [18] Somsen GA, Van Rijn C, Kooij S, Bem RA, Bonn D. Small droplet aerosols in poorly ventilated spaces and SARS-Cov-2 transmission. *Lancet Respir Med*. July 2020. 8(7): 658-659
- [19] K. Team, Contact Transmission of COVID-19 in South Korea: Novel investigation techniques for tracing contacts. *Osong Public Health Res. Prospect.*, Vol. 11, pp. 60-63, Feb. 2020
- [20] [Online] Digital Contact Tracing Tools, 2020. Available at <https://www.cdc.gov/coronavirus/2019-ncov/downloads/digital-contact-tracing.pdf>
- [21] [Online] Digital tools for COVID-19 contact tracing. Annex: contact tracing in the context of COVID-19, Jun 2, 2020. Available at file://winfsd4/users\$/axr249/Downloads/WHO-2019-nCoV-Contact_Tracing-Tools_Annex-2020.1-eng%20(1).pdf
- [22] [Online] Tracetogther, 2020. Available at <https://www.tracetogther.gov.sg>
- [23] [Online] HealthyTogether: Free COVID-19 symptom checking solution for Utah businesses, schools, and select partners, 2020. Available at <https://www.healthytogether.io/>
- [24] [Online] Safe2: Your COVID-19 exposure warning system for smarter social distancing, 2020. Available at <https://safe2.org/>
- [25] [Online] CoEpi: Community Epidemiology in Action, 2020. Available at <https://www.coepi.org/>
- [26] [Online] NOVID: Stop COVID-19 with NOVID, 2020, Available at <https://www.novid.org/>
- [27] [Online] Care-19 Alert, 2020. Available at <https://ndresponse.gov/covid-19-resources/care19>
- [28] M. Ammar, G. Russello, and B. Crispo, Internet of Things: A survey on the security of IoT frameworks, *J. Information Security Application*, VI. 38, pp.8-27, Feb. 2018
- [29] A. W. Dent, A survey of certificateless encryption schemes and security models, *International Journal of Information Security*, Vol. 7, no. 5, pp. 349-377, Oct. 2008
- [30] A. Karati, S. H. Islam and M. Karupiah, Provably secure and lightweight certificateless signature scheme for IoT environments, *IEEE Transaction Ind. Informatics*, vol. 14, no. 8, pp. 3701-3711, Aug. 2018
- [31] Hien Thi Thu Truong, Miguel Almeida, Ghassan Karame, Claudio Soriente, Towards Secure and Decentralized Sharing of IoT Data, *IEEE International Conference on Blockchain*, 2019, Atlanta, GA

International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 11, Issue 03, March 2021)

- [32] Hao Guo, Wanxin Li, Mark Nejad, Chien-Chung Shen, Access Control for Electronic Health Records with Hybrid Blockchain-Edge Architecture, IEEE International Conference on Blockchain, 2019, Atlanta, GA
- [33] R. Aswini, K. Kiruba, College Fees Transaction Using Hash Functions of Blockchain Model, IEEE International Conference on System, Computation, Automation, and Networking (ICSCAN), 2019, Pondicherry, India, India
- [34] Rui Zhang, Rui Xue, and Ling Liu, Security and Privacy on Blockchain, ACM Computer Survey 1, 1, Article 1, 2019
- [35] Susha Surendran, Amira Nassef, Babak D. Beheshti, A survey of cryptographic algorithms for IoT device, IEEE Long Island Systems, Applications, and Technology Conference (LISAT), 2018, Farmingdale, NY, USA
- [36] Yonghao Gu, Weiming Wu, A light-weight mutual authentication protocol for ISO 18000-6B standard RFID system, IEEE International Conference on Communications Technology and Applications, Beijing, China
- [37] Ruhul Amin, G P Biswas, Anonymity preserving secure hash function based authentication scheme for consumer USB mass storage device, Proceeding of 2015, Third International Conference on Computer, Communication, Control and Information Technology (C3IT), Hooghly, India
- [38] Yu-Jia Chen, Li-Chun Wang, Privacy Protection for Internet of Drones: A Network Coding Approach
- [39] [Online] WHO, Coronavirus disease (COVID-19) Situation Report-203, available at https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200810-covid-19-sitrep-203.pdf?sfvrsn=aa050308_4