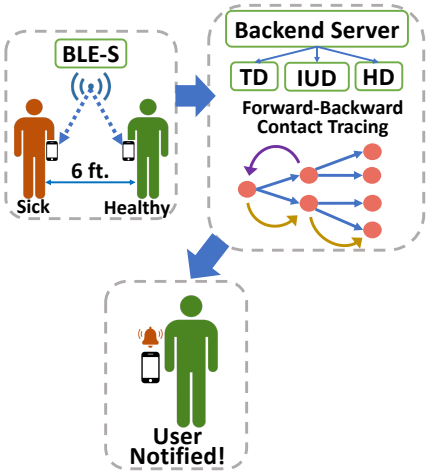
Enhancing Infectious Disease Outbreak Surveillance via Bidirectional Contact Tracing

Akshay Madan, Gaurang Bansal, *Member, IEEE*, Vinay Chamola, *Senior Member, IEEE*, David Tipper, *Senior Member, IEEE*

***Abstract*— Contact tracing remains essential in mitigating the spread of pandemics (including COVID-19). Specifically, backward contact tracing helps find superspreaders and hidden chains of transmission from asymptomatically infected users. However, most literature proposing contact tracing frameworks and apps deployed by various countries do not attempt backward contact tracing. In this work, we present a novel approach for Bidirectional contact tracing. The proposed approach works using Bluetooth low-energy sensors that detect the presence of users in a vicinity and inform a central backend server of user presence. By fixing BLE-S in buildings, the proposed framework can trace the contacts resulting from contamination of a location (*indirect contacts*). We present two algorithms using which the proposed framework can trace forward and backward contacts. Using a simulation, we also track the spread of infection among different 'Generations' of the infected and the impact of backward tracing on preventing the spread across generations. We observe the effect of critical epidemiological parameters such as the reproduction number (R) and the overdispersion parameter (k) specifically on backward contact tracing efficiency.**

***Index Terms*— Global Pandemic, COVID-19, Indirect Contact Tracing, Fomite Transfer, Bluetooth Low Energy (BLE), Backward Contact Tracing**



I. Introduction

Contact Tracing (CT), as a public health tool, can identify individuals who have been in contact with users infected with communicable diseases (including COVID-19) in the recent past. Traditional contact tracing, such as manual efforts, depends on an individual's memory and willingness to reveal their location and contact information. In contrast, Digital CT has the potential to be more effective in the early detection of infectious cases and thus allows for timely medical intervention. It can also help identify individuals who infect a large group of people, commonly known as superspreaders. Since smartphones are already prevalent today, several digital contact tracing applications have been developed using them. Some works use GPS communication for contact tracing (India's Aarogya Setu App), while others use Wi-Fi for user localization and contact tracing [1]. However, due to its low energy requirements, BLE or Bluetooth Low Energy has been a popular technology for smartphone-based contact tracing in many previous works [2], [3]. The most common approach for the same is the broadcast of BLE packets using apps installed on smartphones to other app users in the vicinity and thus register contacts.

Most of the literature proposing contact tracing frameworks and the contact tracing apps deployed by various countries focuses on forward contact tracing and ignoring tracing the contact source. **Backward Contact Tracing** traces the asymptomatic or undetected transmission sources for a given user and thus finds additional chains of transmission originating from them. Due to the overdispersive nature of diseases like COVID-19, users traced by backward CT are also more likely to be superspreaders of disease, boosting the overall infected detection rate. Overdispersion is explained in Section V. Another issue in many contact tracing apps is user privacy [4]. Also, these approaches suffer from lower contact detection rates due to a lack of control over the BLE range. Continuous broadcasting of Bluetooth packets raises issues of rapid battery drain. Due to such reasons, most of the contact tracing attempts in the recent past have failed to reach the minimum app adoption rate required to make them effective [5], [6].

This work proposes a novel contact tracing approach for forward and backward contact tracing using Bluetooth low energy sensors or BLE-S. The proposed framework can detect indirect contacts like transmission by touching surfaces where the contaminants can stay from minutes to several hours [7]. We specifically target environments like college campuses, corporate campuses, prisons, etc., where app usage can be incentivized or made mandatory. This ensures the required

Akshay Madan *(Corresponding author)* (akm88@pitt.edu) & **Dr. David Tipper** (dtipper@pitt.edu) are with the Dept. of Informatics and Networked Systems, University of Pittsburgh, US. **Gaurang Bansal** (gaurang@u.nus.edu) is with the, Dept. of Electrical and Computer Engineering, National University of Singapore, Singapore. **Dr. Vinay Chamola** (vinay.chamola@pilani.bits-pilani.ac.in) is with the, Dept. of Electrical & Electronics Engineering, Birla Institute of Technology and Science, Pilani, India.

TABLE I
NOTATIONS TABLE

NOTATION	MEANING
CDC	Center for Disease Control and Prevention
ξ	Time duration threshold for close contact (in minutes)
δ	Distance threshold for close contact
$\mathcal{T}_m$	Threshold for contamination
$\mathcal{T}_w$	Duration of contamination
θ	Contact Duration
θ_{D}	Direct Contact Duration
θ_{I}	Indirect Contact Duration
X	Set of all known infected users
BLE-S	Bluetooth Low Energy Sensors
BS	Backend Server
BUID	BLE-S Unique Identifier
B	BLE Sensor ID example ID
TD	Token Database
IDT	Infected Indirect Dummy Token
IUD	Infected User Database
HUD	Healthy User Database
TS	Timestamp
UUID	User Unique Identifier
U	User example ID
Y	Infected User
PI	Pre-infection period of index user
f	No. of infectees of an infected user
$p(f)$	Probability Distribution Function of f
Q	Infected Generation
R	Reproduction number
k	Overdispersion parameter
ρ	Coefficient of variation of f
$E(x)$	Expected value of variable x
w	Relative reduction in R of Q_2 due to quarantine
s	Accuracy of backward contact tracing
α	Accuracy of forward contact tracing (direct & indirect)
t	Accuracy and usage probability of COVID test
Π	No. of Q_3 cases prevented by contact tracing
Ψ	Proportion of Q_3 cases prevented

minimum adoption rate. The BLE-S detects the presence of participating nearby user phones, and the user devices send BLE *tokens* to the BLE-S. The tokens are transmitted from the BLE-S to a Backend Server (BS). The BS uses the token information to perform forward and backward tracing of contacts of both types - direct (transmission via touching, conversing directly with infected) and indirect (transmission via contaminated surfaces - fomites or viral particles in air). The contributions of our work can be summarized as -

- Proposing novel technique for bidirectional contact tracing using Bluetooth beacon devices.
- Proposing two algorithms for forward and backward contact tracing.
- Evaluating the impact on preventing the spread of infectious diseases like COVID-19 by applying an analytical model to the proposed contact tracing technique.
- Numerical results are given to illustrate the impact of the proposed technique on the prevention of disease in later generations and compare with two existing CT works.

II. RELATED WORK

The work in [8] lists and compares several works in contact tracing. In [9], the authors propose EPIC, a framework that uses received signal strength indicators (RSSI) values of Wi-Fi and Bluetooth from neighboring wireless devices. Further, the received data is encrypted and sent to the server for contact tracing. They perform privacy-aware contact tracing as the user data is processed in the anonymized domain to obtain the contact information. Since it relies on RSSI values, the technique requires regular calibration as the accuracy is environment dependent.

In [10], the authors propose PC3T, which uses wearable-based proximity detection and machine learning for contact tracing. Their technique has high accuracy, but they do not consider any physical obstacles in their experiments. This leads to inaccuracies in real-world results due to complex signal propagation in indoor environments. In [11], the authors propose LTESafe, where the CSI data from LTE networks is used to estimate the distance between user devices. The CSI data cannot detect altitude, which may lead to false positives with people on different floors. We compare the results of the proposed work with the above two works - LTESafe [11] and PC3T [10] in Section VI. In [4], the authors designed a device that is fixed in public areas for detecting contacts but can not detect indirect contacts. None of the above techniques take indirect contacts into account as opposed to the proposed technique. However, according to [12], indirect contacts such as fomite transfer may have contributed to 25% of the COVID-related deaths in the UK during the post-lockdown period in 2020. In [13], the authors use absolute locations of the user via GPS and thus detect indirect contacts. However, GPS suffers from inaccuracies in dense urban and indoor environments where satellite signals are obstructed [14]. The proposed work uses local Bluetooth-based interactions and is thus more accurate on smaller scale.

According to [15], [16], backward contact tracing can be highly effective compared to a forward-only approach due to the over-dispersed nature of COVID-19. This is because a traced source of COVID-19 is highly likely to have infected other individuals as well [17]. In [18], the authors analytically predict effects of backward contact tracing and find a dramatic 2-3 times increase in preventing the spread of COVID-19.

III. PROPOSED SYSTEM MODEL

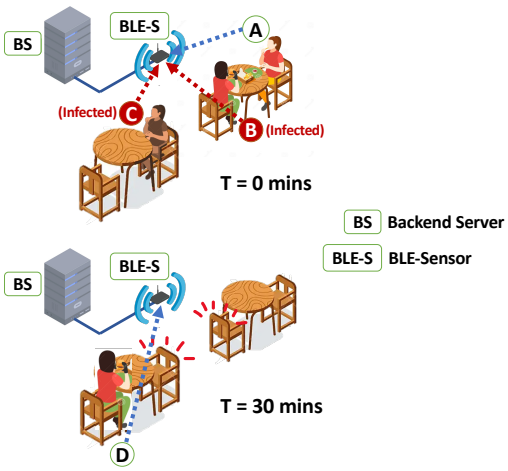


Fig. 1. Example scenario of a restaurant where four users A-D enter and leave at different times. Users' smartphones transmit data to nearby BLE-S connected to the BS. Infected users (A & D) can be detected using the proposed system.

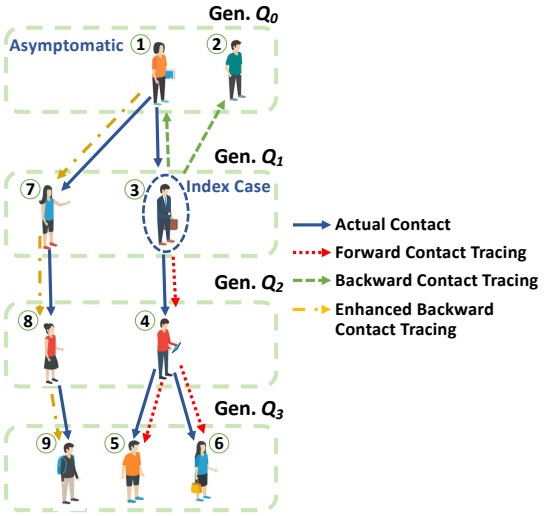


Fig. 2. The spread of COVID-19 among different generations and contact tracing between them

We now describe how the proposed contact tracing system works. Users who volunteer for contact tracing must download the app on their smartphones (**User Equipment** or **UE**) and register with the app. These can be smartphones or devices like a BLE-based keyfob or a smart wearable. Registration assigns every user with a *User Unique Identifier* (UUID) that can not be used to infer any information about the user. This ID also ensures correct contact tracing and that impersonation of an individual is not possible. It can be generated by the server and can include the one-way hash of the user data and random noise to maintain uniqueness and ephemerality. A copy of the UUID is sent to the BS upon every user's registration. The **BLE Sensors (BLE-S)** are installed in buildings. Table I lists notations like UUID and BLE-S adopted within this paper. This work can be targeted to corporate campuses, educational campuses, prisons, etc. This ensures a high adoption rate but sources of infection outside such premises will go undetected. The UE contacts the BLE-S installed in buildings to register contacts. BLE-S are Bluetooth low-energy transceivers that receive location and UUID from nearby users (Fig.1). Upon reception, they store the tokens to detect contacts between users. Directional antennas are used in BLE-S to create a targeted range of Bluetooth with a diameter of 6ft.; thus, two UEs discover the same BLE-S only if they are within 6ft. of each other or less and not otherwise. Such control over the Bluetooth range is impossible in common smartphone peer-peer contact tracing apps. Every BLE-S is assigned a *BLE-Sensor Unique Identifier* (or BUID). Since this ID uniquely identifies the fixed BLE-S, thus it is synonymous with the location of the BLE-S and can also be used to discretely link a token with a relative location. Since these contain sensitive information these IDs are randomly generated and regularly updated. A UE senses if there is a BLE-S in their vicinity using BLE. Once detected, the UE transmits their information in the form of a *token* to the BLE-S at a specified rate as long as they are in the vicinity (Fig.1 - dotted arrows from users to BLE-S). The token consists of the user's UUID and a timestamp.

The BLE-S also connects with the BS. The BS consists of 3 databases. The first database maintains the record of all the tokens received from the users (Token Database or TD) and whether they are infected. The second database accounts for the different registered users, their identifiers, and their total contact time with infected individuals (Healthy User Database HUD). The third database maintains a list of infected users with their user identifiers (Infected User Database or IUD). Given how sensitive the information stored on this server is, it is preferred that the government health authorities maintain the BS. The BLE-S periodically releases all the tokens it has stored in its memory along with its identifier (BUID) to the BS (Fig.1). A token from User Y, received by BLE-S with BUID B at timestamp TS, can be written as $\langle \text{TS}, Y, B \rangle$. The BS receives records from every installed BLE-S about all the users and stores it in the BS-TD. When a user is diagnosed with COVID-19 they are requested to upload the information directly to the BS via the internet, which is stored in BS-IUD. Thus, all the tokens associated with that UUID are marked as infected in the BS database.

With the above information, the BS tries to evaluate healthy users' direct and indirect contacts. Center for Disease Control and Prevention (CDC) defines a *direct close contact* as spending more than $\xi = 15\ mins$ within $\delta = 6ft.$ of infected [19]. As an example, Fig. 1 is a restaurant scenario with the proposed technique. The upper scene and lower scene are 30 minutes apart. In Fig.1, user A experiences direct contact with infected user B. Thus, for each infected token (Fig.1 A-B), the BS searches for all BS-TD tokens with the same BUID and same timestamp. It then adds the contact duration related to that infected token of the corresponding user in the BS-HUD. We assume that an infected user contaminates a place for a window of $\mathcal{T}_w$ beginning from when they leave that place if they spend more than $\mathcal{T}_m$ time there. These values are configurable as per the latest epidemiological research. The phone-phone apps are not capable of detecting such contacts. In Fig.1, user D experiences indirect contact with infected user C even after C has vacated the seat, provided the seat used by C was not cleaned. To detect indirect contacts, the BS notes the time the infected user has spent around a BLE-S using its tokens in the TD. If it is greater than $\mathcal{T}_m$, then the BS looks for tokens in the TD for healthy users within the $\mathcal{T}_w$ window after the infected user has left the vicinity. A corresponding number of contact duration are added for respective healthy users in the BS-HUD. Additionally, the BS periodically checks the HUD for individuals whose total contact duration (direct + indirect) has exceeded the threshold ξ. It informs such healthy users of their possible infection and suggests measures such as testing and isolation.

We call a subset of users part of a **Generation** if they are the same number of transmissions after the source case. Thus, in Fig. 2, user 1 is part of Generation 0 (Q_0) (a.k.a. the source case); the index case and users (3 & 7) part of Generation 1 (Q_1); users 4 & 8 part of Generation 2 (Q_2); and users 5, 6 & 9 part of Generation 3 (Q_3). Fig.2 shows the index case (user 3) (Generation Q_1). So, forward tracing (red dotted arrows) identifies users contacted by index case, i.e., user 4 (Generation Q_2). In Fig.2, backward tracing (green dashed arrows) identifies the source of infection (Generation Q_0) for

Algorithm 1 Forward Contact Tracing - Direct & Indirect

1: **Input:** BS-TD, BS-IUD, $\mathcal{T}_m$, $\mathcal{T}_w$
2: **Output:** BS-HUD
3: $C = \{\}$ ▷ tokens already accounted for
4: IDT $= \{\}$ ▷ Infected Indirect Dummy Token
5: **for** $Y \in$ {Infected Users} **do**
6: **// Direct Contact Tracing**
7: **for** token $T_Y = \langle \text{TS}, B, Y \rangle \in \{Y$'s Tokens$\}$ **do**
8: **if** $\langle \text{TS}, B, (\text{UUID} = \text{Any}) \rangle \notin C$ **then**
9: C.**append**$(\langle \text{TS}, B, (\text{UUID} = \text{Any}) \rangle)$
10: $L = \langle \text{TS}, B, U \rangle \in$ BS-TD, $\forall\, B, U \neq Y$
11: **for** token $T \in L$ **do**
12: BS-HUD$(U)_{\text{direct}}$.**append**(T)
13: **// Indirect Contact Tracing**
14: Array A = Contiguous array of Y's tokens
15: for A, let BUID = B_A
16: **if** duration$(A) > \mathcal{T}_m (= 3\ mins)$ **then**
17: A_e = last timestamp for A
18: IDT.**append**$\{\langle A_e + 1, B_A, Y \rangle, \langle A_e + 2, B_A, Y \rangle, \cdots, \langle A_e + \mathcal{T}_w, B_A, Y \rangle\}$ ▷ $\mathcal{T}_w = 15\ mins$
19: **for** token $T_{\text{IDT}} = \langle \text{TS}, B_A, Y \rangle \in$ IDT **do**
20: **if** $\langle \text{TS}, B_A, (\text{UUID} = \text{Any}) \rangle \notin C$ **then**
21: C.**append**$(\langle \text{TS}, B_A, (\text{UUID} = \text{Any}) \rangle)$
22: $L = \langle \text{TS}, B_A, U \rangle \in$ BS-TD, $\forall\, B, U \neq Y$
23: **for** token $T \in L$ **do**
24: BS-HUD$(U)_{\text{indirect}}$.**append**(T)

the index case user 3, that is, users 1 & 2.

To detect backward contacts in the proposed framework, the BS looks at the tokens in the BS-TD for the infected index user (Fig.2 user 3) in the pre-infection period of the user. For COVID-19, the symptoms manifest between 2-14 days after exposure [20]; we assume this to be the pre-infection period of the index user. The BS looks for tokens of other healthy users with the same BUID as the index user and the same timestamp - users 1 & 2 during this period. If in Fig. 2 user 1 has sufficient tokens with the index user 3, then user 1 may be infected and a possible source of infection for the index case user 3. We use enhanced backward tracing by applying forward tracing on the Generation Q_0 user in the BS-TD. If user 1 in Fig.2 is identified as the source of infection for index case user 3, then we can apply forward tracing on user 1 to find another chain of possibly infected users, i.e., 7, 8, & 9.

IV. Forward & Backward Tracing Algorithms

Now, we describe the forward and backward contact tracing algorithms. We assume that the healthy users contact disease only from infected who are participating in contact tracing. Another assumption we make is that infected users always volunteer to reveal their diagnosis. In Algorithm 1, we show the forward contact tracing algorithm for both direct and indirect contacts. First, we consider every token T_Y of every known infected user Y (lines 3 and 5). Then, in line 8, we search for more tokens in the BS-TD such that the timestamp is the same as the considered token (TS$_Y$) and also the BUID (B) and store into a temporary variable L.

$$L = \langle \text{TS}, B, U \rangle \in \text{BS-TD}, \forall\, B, U \neq Y$$

In lines 6 and 7, we ensure that the combination of timestamp TS$_Y$ and the BUID = B are never repeated once identified as infected; thus, we create an array C for bookkeeping.

If $\langle \text{TS}, B, (\text{UUID} = \text{Any}) \rangle \notin C$ **then,**
C.**append**$(\langle \text{TS}, B, (\text{UUID} = \text{Any}) \rangle)$

In line 10, we add the direct contact duration incurred by tokens in L to corresponding healthy users in the BS-HUD.

BS-HUD$(U)_{\text{direct}}$.**append**(T)

After this, we move to indirect contact detection. In line 12, we look for a contiguous array A of tokens with the same BUID (B_A) and UUID as the infected user Y. If this array of tokens A has a timestamp duration greater than $\mathcal{T}_m (= 3\ mins)$, then the user Y has spent enough time to cause indirect contacts. Thus, a time window of $\mathcal{T}_w (= 15\ mins)$ is infected around BLE-S with BUID = B_A beginning right after the array's last timestamp A. Since user Y has already left, there are no tokens during that time around the BLE-S B_A to match healthy user tokens with. Therefore, we create an array of dummy tokens to mark the contamination of the place, *Infected Indirect Dummy Token*, or IDT, for the duration of this window in line 15.

$$\text{IDT}.\textbf{append}\{\langle A_e + 1, B_A, Y \rangle, \langle A_e + 2, B_A, Y \rangle, \cdots, \langle A_e + \mathcal{T}_w, B_A, Y \rangle\}$$

After this, we find tokens in BS-TD with the same BUID and timestamp as tokens in IDT (line 19).

$$L = \langle \text{TS}, B_A, U \rangle \in \text{BS-TD}, \forall\, B, U \neq Y$$

Finally, we add the corresponding duration of indirect contact to corresponding healthy users in BS-HUD (line 21).

BS-HUD$(U)_{\text{indirect}}$.**append**(T)

In Algorithm 2, we discuss how to perform backward contact tracing using the proposed system for direct and indirect contacts. For an index user, Y, who is now infected, we assume the Pre-infection period (PI$_Y$) as,

$$\text{PI}_Y = [\text{TS}_{Y-\text{symp}} - 14\text{days},\ \text{TS}_{Y-\text{symp}} - 2\text{days}].$$

We divide the dataset of tokens into those received by the infected user Y (set T_1) and by others (T_2) during this pre-infection period. T_2 can be filtered wisely, picking only those users in the same city neighborhood as user Y and thus likely to have been the source of infection for Y.

$$T_1 = \langle \text{TS}, B, U = Y \rangle \in \text{BS-TD}\ \forall\ \text{TS} \in \text{PI}_Y$$
$$T_2 = \langle \text{TS}, B, U \neq Y \rangle \in \text{BS-TD}\ \forall\ \text{TS} \in \text{PI}_Y$$

For every user U in T_2, we find the total contact duration, Θ, between U & Y (described in detail in Algorithm 1). If that is greater than the time threshold for making contacts defined by the CDC, i.e., $\xi\ (= 15\ mins)$, then it is a candidate for the source of infection for index case, Y (lines 9-11).

$\Theta =$ Total direct contact duration b/w U & Y
If $\Theta > \xi$ **then** CBC.**append**(U)

However, the index user may have also gotten the infection via indirect contact and our technique can detect this (lines 13-20). We look for contiguous arrays of tokens in T_2 for

Algorithm 2 Backward Contact Tracing - Direct & Indirect

1: **Input:** PI_Y, BS-TD, BS-IUD, $\mathcal{T}_m$, $\mathcal{T}_w$
2: **Output:** BS-HUD
3: CBC = {} ▷ candidate backward contacts
4: **for** $Y \in$ {Infected Users} **do** ▷ index case user
5: // PI_Y - Pre-infection period of index case user Y
6: $\text{PI}_Y = [\text{TS}_{Y-\text{symp}} - 14\text{days},\ \text{TS}_{Y-\text{symp}} - 2\text{days}]$
7: $T_1 = \langle \text{TS}, B, U = Y \rangle \in \text{BS-TD} \,\forall\, \text{TS} \in \text{PI}_Y$
8: $T_2 = \langle \text{TS}, B, U \neq Y \rangle \in \text{BS-TD} \,\forall\, \text{TS} \in \text{PI}_Y$
9: **for** User $U \in T_2$ **do**
10: **// Direct Contact Tracing**
11: Θ = Total direct contact duration b/w U & Y
12: **if** $\Theta > \xi$ **then** ▷ ξ is contact duration threshold (= 15 *mins*)
13: CBC.**append**(U)
14: **// Indirect Contact Tracing**
15: Array A = contiguous token array in T_2
16: for A, let BUID = B_A
17: **if** Duration(A) > $\mathcal{T}_m$(= 3 *mins*) **then**
18: A_e = last timestamp(A)
19: IDT = $\{\langle A_e{+}1, B_A, U\rangle, \langle A_e{+}2, B_A, U\rangle, \cdots, \langle A_e{+}\mathcal{T}_w, B_A, U\rangle\}$ ▷ ($\mathcal{T}_w$ = 15 *mins*)
20: T_{CBC} = IDT $\cap$ T_1
21: **if** $T_{\text{CBC}} > \xi$ **then**
22: CBC.**append**(U)

every user. If we find such an array with a duration greater than the threshold for indirect contacts $\mathcal{T}_m(= 3\ mins)$, then we create an Infected Indirect Dummy Token, IDT, same as in Algorithm 1.

$$\textbf{If } \text{Duration}(A) > \mathcal{T}_m \textbf{ then}$$
$$\text{IDT} = \{\langle A_e{+}1, B_A, U\rangle, \langle A_e{+}2, B_A, U\rangle, \cdots, \langle A_e{+}\mathcal{T}_w, B_A, U\rangle\}$$

Finally, similar to lines 10-11, we find common tokens between IDT and T_1, and if it is greater than the time threshold for making contacts $\xi(= 15\ mins)$, it is another candidate for the source of infection for index case, Y. Finally, candidate sources from both kinds of contacts are informed of being sick asymptomatically if not already known to be infected.

V. Simulation Setup

We use a branching process to simulate the effectiveness of combining 'backward' tracing with traditional 'forward' tracing when dealing with the presence of overdispersion in the COVID-19 transmission based on the works - [15], [18], [21]. ***Overdispersion*** is a phenomenon where the distributions in the number of infected has a much higher variability instead of a homogenous spread. This gives rise to concepts like *superspreaders* and *hotspots* where a smaller population is responsible for a disproportionate number of transmissions.

Through our simulation, we examine the effects of our approach, which involves both forward and backward tracing methods. We assume a tree-like structure for the transmission of disease (e.g., Fig. 2, solid blue arrows). We assume that Generation Q_0 user 1 can be found using backward tracing from the known-infected index user. Also, users in Q_1, including the index case user, are known to be infected through symptomatic diagnosis. Some of users in Q_2 are also exposed to the disease but are well informed before becoming infectious through contact tracing. Based on Generation Q_2 users' willingness to comply with quarantine, we study the effect of contact tracing in saving the users in Generation Q_3 from getting infected. Due to the overdispersive nature of COVID-19-like diseases' successor distribution, the number of successors of any random individual is expectantly less than those identified by backward tracing (source case) as the latter are already known to have at least one successor (index case). Thus, the source case is likelier to infect other nodes than the index case. Let $p(f)$ represent the probability mass function of f, the number of successor transmissions originating from any diagnosed case. Thus, the probability distribution of several successors of the identified source case (Q_0) is $p(f|Q_0)$.

$$p(f|Q_0) = \frac{f \cdot p(f)}{\mathbf{E}(f)}. \tag{1}$$

where $\mathbf{E}(f) = \sum_{f=0}^{\infty} f \cdot p(f)$, is the expectation of f. The expected number of Q_1 cases that can be identified through backward tracing (inclusive of the index case) is expressed as:

$$\mathbf{E}(f|Q_0) = \sum_{f=0}^{\infty} f \cdot p(f|Q_0) \tag{2}$$
$$= \frac{\mu^2 + \sigma^2}{\mu} \tag{3}$$

Here, μ is the mean of f and σ^2 is the variance of f. We assume that $p(f)$ follows a negative binomial distribution with an overdispersion parameter k, a common assumption in previous literature [22]. Thus, according to [23], the variance of f with a negative binomial distribution follows,

$$\sigma^2 = \mu + \frac{\mu^2}{k} \tag{4}$$

The Reproduction number, R (also denoted by R_0), is commonly defined as the average number of users infected by any user, which is numerically the same as μ. This implies that the diseases that are transmitted easily and also by fomite-based contacts will have a higher value of R. Analyzing the R value is important for risk assessment, public health planning, and policy decision-making. Thus, the expected number of Q_1 cases identified by backward tracing are:

$$\mathbf{E}(f|Q_0) = 1 + R \cdot \left(1 + \frac{1}{k}\right) \tag{5}$$

As explained before overdispersion leads to a disproportionate spreading of the virus. And, a smaller value of k indicates a higher degree of overdispersion. This gives rise to concepts like superspreaders who are more likely to spread the virus than the general population. COVID-19 is known to have a small $k \in [0.1, 0.5]$. It is noteworthy that $[\mathbf{E}(f|Q_0) = 1 + R(1 + \frac{1}{k})] > [\mathbf{E}(f) = R]$, which proves what was stated earlier, that due to overdispersion in f, the expected number of successors of a random individual is expectantly less than those identified by backward tracing. In the simulation, we vary R & k and predict the average number of Q_1 users identified by backward tracing ($\mathbf{E}(f|Q_0)$) and the probability it is larger than 5, 10, 15, etc.

We measure the efficiency of the proposed method, defined as the relative reduction in the number of Generation 3 or Q_3 cases. Due to timely action by Q_2 cases such as isolation, etc., several Q_3 cases are averted (A_3). We estimate A_3 as,

$$A_3 = \sum_{f_0,f_1,f_2=0}^{\infty} (f_0 \cdot p(f_0|Q_0))(f_1 \cdot p(f_1))(f_2 \cdot p(f_2)) = R^2 \cdot \left(1 + R \cdot \left(1 + \frac{1}{k}\right)\right). \quad (6)$$

First, we ignore the backward tracing and only focus on forward tracing. The total number of Q_1 cases excluding the index case is $\mathbf{E}(f|Q_0) - 1 = R(1 + \frac{1}{k})$. In the absence of backward tracing, the source case remains unknown. Thus, only cases in Q_1 for whom further (forward-only) contact tracing will occur are those detected through random population testing independent of contact tracing. We assume t to be the fraction of cases detected through independent disease testing. Thus, the total cases detected in Q_1 (including index case) are $1 + R \cdot t(1 + \frac{1}{k})$. The number of Q_2 cases generated by all the Q_1 cases are $R(1 + R \cdot t(1 + \frac{1}{k}))$, assuming R is the average reproduction number. We assume the accuracy of the forward contact tracing to be α, then only $R \cdot \alpha(1 + R \cdot t(1 + \frac{1}{k}))$ users will be detected in Generation Q_2 by the contact tracing. α represents the accuracy of detecting both kinds of contacts, direct & indirect. A higher value of α means a more effective contact tracing technique in place which requires a sufficient proportion of the population to participate in the contact tracing. Our previous work [3] also detected both direct and indirect contacts and used a beacon infrastructure to detect contacts. We use the same accuracy value derived from that work in our simulation in Section VI as an approximation.

We assume that due to isolation, individuals belonging to Q_2 have a reduced reproduction rate, $w * R$ $w \in [0, 1]$. Thus, the number of Q_3 cases that remain uninfected benefiting from forward-only contact tracing are:

$$\Pi_F = (1 - w) \cdot R\left[R \cdot \alpha\left[1 + R \cdot t\left(1 + \frac{1}{k}\right)\right]\right]. \quad (7)$$

In combined forward & backward tracing also, the total number of Q_1 cases are $R(1+\frac{1}{k})$. We assume that the probability of identifying the source case from the index case through contact tracing is s. This is essentially the accuracy of the backward contact tracing (Algorithm 2). A higher s would imply the better efficiency of the backward contact tracing method. Thus, the effective proportion of Q_1 cases identified by applying forward tracing on the source case (Q_0) is $s\alpha$. Thus, the proportion detected either through backward tracing or independent testing is $(1 - (1 - t)(1 - s\alpha))$. Following similar logic as the forward tracing case, the total number of cases identified in Q_1, given only the index case, are $1 + R[1 - (1-t)(1-s\alpha)][1 + \frac{1}{k}]$. And the identified Q_2 cases are $R \cdot \alpha\left[1+R[1-(1-t)(1-s\alpha)][1+\frac{1}{k}]\right]$. Finally, the number of Q_3 cases that were averted because of forward and backward contact tracing are:

$$\Pi_{FB} = (1-w) \cdot R\left[R \cdot \alpha\left[1+R[1-(1-t)(1-s\alpha)][1+\frac{1}{k}]\right]\right] \quad (8)$$

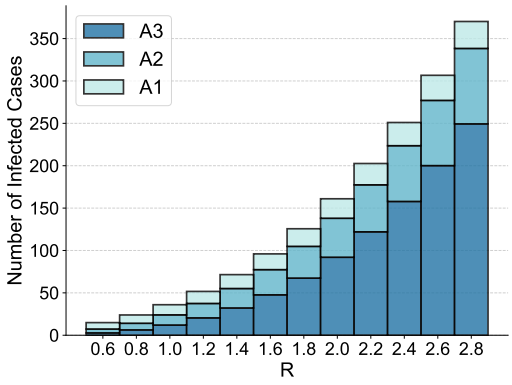


Fig. 3. Number of cases, A_1, A_2, & A_3, in Generations Q_1, Q_2, & Q_3 vs R

Thus, the efficiency of contact tracing in the case of forward-only tracing and forward-backward tracing is -

$$\Psi_F = \Pi_F / A_3 \quad (9)$$

$$\Psi_{FB} = \Pi_{FB} / A_3. \quad (10)$$

VI. Results & Discussion

Drawing upon the mathematical formulations presented in Section V, we developed a Python-based simulation model. Our simulator extends the framework introduced in [18]. While the latter provides a general analysis, we adapt it to evaluate the proposed method specifically along with two other existing techniques. The purpose of the simulator is to study the efficiency of the proposed forward and backward contact tracing technique. In Fig. 3, we study the no. of cases in different generations $A_1 - A_3$ w.r.t. R. We also observe the contribution of different generations. The number of cases increases much more rapidly in the case of Q_3 than Q_2 or Q_1. At the largest value of R considered ($R = 2.8$), Q_3 contributes to around 67% of the total cases.

In our simulations, we assume a single source case (Q_0), which produces various successors in three subsequent generations ($Q_1 - Q_3$). We observe the efficiency contact tracing has on the prevention of COVID-19 in Generation Q_3 (Eqns. 7 and 8). Alongside, we study the effects of parameters like α, R, k, etc. on the same. R has an estimated range of $\{1.2, 2.5\}$ for COVID-19, thus for a more inclusive range we assume $R \in \{0.8, 1.2, 2.5\}$. In Fig. 4, we show the proportion of Q_3 cases prevented by forward-only tracing (Ψ_F) and also by forward-backward tracing (Ψ_{FB}), for different values of R. We plot three bands/regions of different colors (see Fig. 4 legend) for different values of R. The upper dashed line for each region has $w = 1$ (no reduction in R due to quarantine by Q_2) whereas the bottom dashed lines denote $w = 0.2$ ($R_{\text{eff}} = 0.2 \times R$). Thus, vertically, the band represents the range of contact tracing efficiency depending upon the different degrees of quarantine in Generation Q_2. A fundamental insight we gain from this figure is that Ψ_F & Ψ_{FB} rise as R increases. A higher value of R means a higher number of infected users. This finding is consistent with conclusions drawn in our earlier study [3], which demonstrated that tracing becomes more efficient as the pool of individuals to be traced grows.

We compare the proposed work to two other existing works - **PC3T** [24] and **LTESafe** [11] (described in Section II).

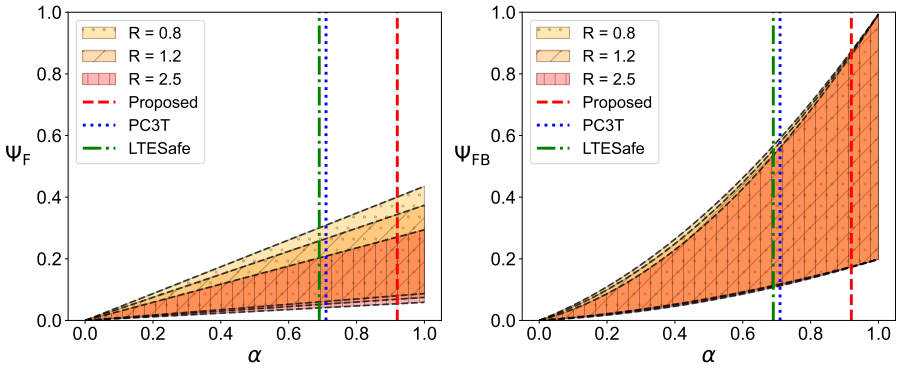


Fig. 4. $\boldsymbol{\Psi_F}$ & $\boldsymbol{\Psi_{FB}}$ vs $\boldsymbol{\alpha}$ for different values of $\boldsymbol{R}$ ($\boldsymbol{s = 0.99, t = 0.2, k = 0.5}$). The dashed lines on top of the colored area correspond to $\boldsymbol{w = 1}$, whereas those at the bottom are for $\boldsymbol{w = 0.2}$, and $\boldsymbol{w}$ varies vertically.

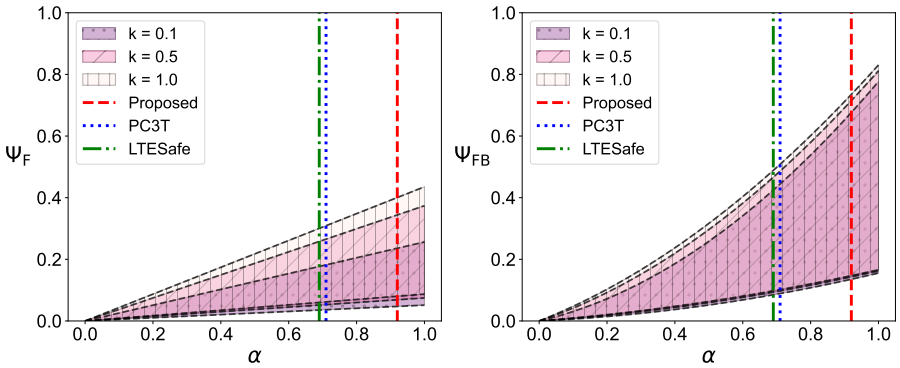


Fig. 5. $\boldsymbol{\Psi_F}$ & $\boldsymbol{\Psi_{FB}}$ vs $\boldsymbol{\alpha}$ for different values of $\boldsymbol{k}$ ($\boldsymbol{R = 1.2, s = 0.7, t = 0.2}$). The dashed lines on top of the colored area correspond to $\boldsymbol{w = 1}$, whereas those at the bottom are for $\boldsymbol{w = 0.2}$, and $\boldsymbol{w}$ varies vertically.

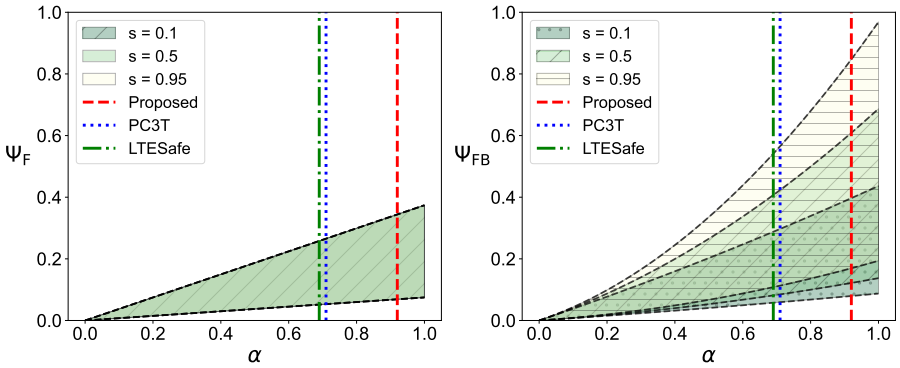


Fig. 6. $\boldsymbol{\Psi_F}$ & $\boldsymbol{\Psi_{FB}}$ vs $\boldsymbol{\alpha}$ for different values of $\boldsymbol{s}$ ($\boldsymbol{R = 1.2, t = 0.2, k = 0.5}$). The dashed lines on top of the colored area correspond to $\boldsymbol{w = 1}$, whereas those at the bottom are for $\boldsymbol{w = 0.2}$, and $\boldsymbol{w}$ varies vertically.

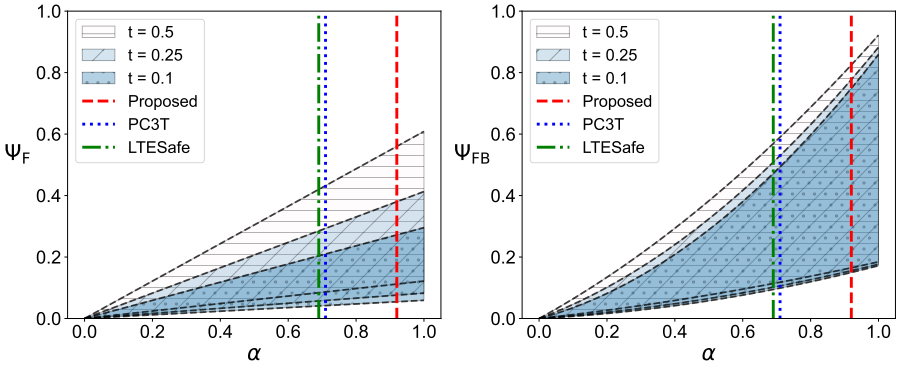


Fig. 7. $\boldsymbol{\Psi_F}$ & $\boldsymbol{\Psi_{FB}}$ vs $\boldsymbol{\alpha}$ for different values of $\boldsymbol{t}$ ($\boldsymbol{R = 1.2, s = 0.8, k = 0.5}$). The dashed line on top of each colored area corresponds to $\boldsymbol{w = 1}$, whereas those at the bottom are for $\boldsymbol{w = 0.2}$, and $\boldsymbol{w}$ increases vertically upwards.

Although there are many works in this area of contact tracing, to show comparison with the proposed method we pick these two state-of-art techniques. Additionally, they are much closer to our work and make for a fair comparison. We have also drawn vertical lines indicating the values of α or the detection rate of our present work ($\alpha = 0.92$, mentioned before in Section V) and also from PC3T [24] (0.942) and LTESafe [11] (0.928). The values of accuracy mentioned in [24] and [11] are only referring to direct contacts (α_d). In reality, there must have also been indirect contacts that were never accounted for. As mentioned in Section II, according to [12], 25% of deaths in the UK during the post-lockdown period were due to fomite transfers (touching surfaces). So, the total transmissions due to fomite transfer and virus-containing particles in the air will be greater than deaths. Thus, the effective α for [24] and [11] is expected to be more than 25%. Thus, an approximation of the effective α for [24] and [11] can be $\alpha \leq 0.75 \times \alpha_d$, or 0.706 & 0.696, respectively since they only consider direct contacts.

Both, Ψ_F & Ψ_{FB}, are better in the case of the proposed method when compared to existing works (PC3T and LTE-Safe) (Fig. 4) due to higher accuracy of BECT. However, the difference is more significant in the case of forward-backward efficiency instead of forward only. Similarly, in Fig. 5, 6 and 7, instead of R we show regions with different values of k, s & t while α and w vary the same as before. In Fig. 5, we observe that forward-only tracing (Ψ_F) bears the impact of varying k or overdispersion parameters. We assume the values of $k \in \{0.1, 0.5, 1\}$ to include the case of lowest overdispersion to maximum. However, this effect is almost neutralized by the addition of backward tracing. Ψ_{FB} is almost independent of variation in k as α increases. On the other hand, in Fig. 6, we observe the opposite effect. There is no impact of s on forward tracing (Ψ_F). This is because s corresponds to the accuracy of identifying the source (backward tracing), so a higher s will not impact forward-only efficiency. But, Ψ_{FB} sees a significant difference with varying values s. We vary the values of $s \in \{0.1, 0.5, 0.95\}$ to account for both reliable and inaccurate backward contact tracing technique. Considering the efficiency of forward-only tracing, we observe that the variable t has the maximum impact on Ψ_F, while it has little effect on Ψ_{FB}. Thus, one can focus on improving the accuracy of independent testing to have higher forward-only efficiency.

VII. Conclusions

In this work, we propose a contact tracing approach for forward and backward contact tracing. The proposed approach uses Bluetooth sensors that detect the presence of users in a vicinity and inform a centralized BS about anonymized user locations. Fixed sensors allow for detecting indirect contacts between users like fomite transfer. The BS can detect forward and backward contacts of a given user. We conducted a simulation to evaluate the impact of the proposed technique on the prevention of the spread of COVID-19 and compared the results with two other existing works. We found improvement in averting COVID-19 cases when compared to existing works. However, this work has a few limitations. The privacy risk to the users are not evaluated given their health data is stored into an external server. We have analyzed the work using simulation and performance analysis of the proposed CT technique in the real world is yet to be done.

References

[1] G. Li, S. Hu, S. Zhong, W. L. Tsui, and S.-H. G. Chan, "Vcontact: Private wifi-based iot contact tracing with virus lifespan," *IEEE IoT Jnl.*, vol. 9, no. 5, pp. 3465–3480, 2021.

[2] D. J. Leith and S. Farrell, "Measurement-based evaluation of google/apple exposure notification api for proximity detection in a light-rail tram," *Plos one*, vol. 15, no. 9, p. e0239943, 2020.

[3] A. Madan, D. Tipper, B. Palanisamy, M. Abdelhakim, and P. Krishnamurthy, "Bect: Beacon-based contact tracing for detecting direct & indirect contacts," in *2022 IEEE 19th Intl. Conf. on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, 2022, pp. 236–242.

[4] P. Tedeschi, K. E. Jeon, J. She, S. Wong, S. Bakiras, and R. Di Pietro, "Privacy-preserving and sustainable contact tracing using batteryless bluetooth low-energy beacons," *IEEE Security & Privacy*, vol. 20, no. 3, pp. 91–100, 2021.

[5] K. Oyibo, K. S. Sahu, A. Oetomo, and P. P. Morita, "Factors influencing the adoption of contact tracing applications: systematic review and recommendations," *Frontiers in Digital Health*, vol. 4, p. 862466, 2022.

[6] K. Kolasa, F. Mazzi *et al.*, "State of the art in adoption of contact tracing apps and recommendations regarding privacy protection and public health: Systematic review," *JMIR mHealth and uHealth*, vol. 9, no. 6, p. e23250, 2021.

[7] "Modes of transmission of virus causing covid-19," 2021, [Online link].

[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] T. Altuwaiyan, M. Hadian, and X. Liang, "Epic: efficient privacy-preserving contact tracing for infection detection," in *2018 IEEE Intl. Conf. on Comm. (ICC)*. IEEE, May 2018, pp. 1–6.

[10] P. C. Ng, P. Spachos, and K. N. Plataniotis, "Covid-19 and your smartphone: Ble-based smart contact tracing," *IEEE Systems Jnl.*, vol. 15, no. 4, pp. 5367–5378, 2021.

[11] F. Yi, Y. Xie, and K. Jamieson, "Cellular-assisted covid-19 contact tracing," in *Proc. of the 2nd Workshop on Deep Learning for Wellbeing Applications Leveraging Mobile Devices and Edge Computing*, 2021, pp. 1–6.

[12] A. Meiksin, "Dynamics of covid-19 transmission including indirect transmission mechanisms: a mathematical analysis," *Epidemiology & Infection*, vol. 148, 2020.

[13] L. Xiong, C. Shahabi, Y. Da, R. Ahuja, V. Hertzberg *et al.*, "React: Real-time contact tracing and risk monitoring using privacy-enhanced mobile tracking," *SIGSPATIAL Special*, vol. 12, no. 2, pp. 3–14, 2020.

[14] P. A. Zandbergen, "Accuracy of iphone locations: A comparison of assisted gps, wifi and cellular positioning," *Transactions in GIS*, vol. 13, pp. 5–25, 2009.

[15] A. Endo, S. Abbott, A. J. Kucharski, S. Funk *et al.*, "Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china," *Wellcome open research*, vol. 5, 2020.

[16] S. Kojaku, L. Hébert-Dufresne, E. Mones, S. Lehmann, and Y.-Y. Ahn, "The effectiveness of backward contact tracing in networks," *Nature physics*, vol. 17, no. 5, pp. 652–658, 2021.

[17] J. Raymenants, C. Geenen, J. Thibaut, K. Nelissen, S. Gorissen, and E. Andre, "Empirical evidence on the efficiency of backward contact tracing in covid-19," *Nature Comm.*, vol. 13, no. 1, p. 4750, 2022.

[18] A. Endo, Q. J. Leclerc, G. M. Knight, G. F. Medley, K. E. Atkins *et al.*, "Implication of backward contact tracing in the presence of overdispersed transmission in covid-19 outbreaks," *Wellcome open research*, vol. 5, 2020.

[19] "Contact tracing for covid-19," 2022, [Online link].

[20] "Symptoms of covid-19," 2022, [Online link].

[21] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, "Superspreading and the effect of individual variation on disease emergence," *Nature*, vol. 438, no. 7066, pp. 355–359, 2005.

[22] S. Coly, M. Garrido, D. Abrial, and A.-F. Yao, "Bayesian hierarchical models for disease mapping applied to contagious pathologies," *PloS One*, vol. 16, no. 1, p. e0222898, 2021.

[23] M. D. Robinson and G. K. Smyth, "Small-sample estimation of negative binomial dispersion, with applications to sage data," *Biostatistics*, vol. 9, no. 2, pp. 321–332, 2008.

[24] P. C. Ng, P. Spachos, S. Gregori, and K. N. Plataniotis, "Epidemic exposure tracking with wearables: A machine learning approach to contact tracing," *IEEE Access*, vol. 10, pp. 14 134–14 148, 2022.