**Features for Data Analytics:**

1. **GPS data:** used only to calculate the probability of getting infection based on statistical features of the region like location category (work, school, shop, hospital, residential area, park), population density. The features are extracted for each location using API calls and NLP on the data
2. **GPS data** is used to map and *trace only the patient path for the past 15 days*. Note the data of other people are stored using (hash-based message authentication code) and cannot be accessed by anyone without the hash code.
3. **Bluetooth: The MAC address of every user phone is hashed in the database.** The Bluetooth of the mobile is used to scan the mac address of other people around them and upload the data as an encrypted package. The mac address can be considered as people and scanning the mac address is equivalent to scanning people. This method also has two advantage over conventional geolocation tracking.
   1. **We do not need to Compare data between people, thus making computation exponentially faster and efficient.**
   2. **In case some people do not download the app and act like outliers. They can be traced in case of contact, as mac address of every device is unique and the device info and owner could be retrieved using Mobile Service Provider. Note the app scan all devices having Bluetooth regardless app is installed or not.**
4. **Human Activity Recognition**: Our API call and algorithms are capable of calculating the motion of the person, i.e. (Still, OnFoot, Biking, Vehicle, Metro). This data can be further analyzed and used for better probability prediction. Thus adding another layer of statistical modelling of predicted region and prediction of the level of contact.

**Each data packets**

|  |  |  |
| --- | --- | --- |
| Time Stamp | HBMA encrypted | Geolocation |
| List of Bluetooth devices near me |
| Activity Performed |

**Method for Maintaining Consistency and Authenticity of data:**

1. **Digital Signature:**  Digital signature is added to the user data while sending to maintain consistency of the data
2. **Doctors/Government verification platform:**  users are not given permission or feature to update their status, i.e. (Healthy, Infected, Quarantined). This provision is only kept for official. The officials or doctors are provided with a portal which allows them to update the status of the patient using the unique key of the patient. The doctor's portal allows them to unlock the patient data for backtracking and analysis **only after the patient shares their Hash code stored in their mobile phone locally. Note the user's data cannot be accessed/decrypted without the hash code used to encrypt the data.**

**The efficiency of the Algorithm**

Designing of the Algorithm is done in such a way that it does not need to analyze every individual's data to make its prediction accurately. In simple words, we just need to analyze the affected patents data to predict every one probability of infection. There is **no need for comparison between data of users.**  As it uses Bluetooth to store people in contact. Simply, if you have come in contact with an infected person then your id would be reflected in the Bluetooth history of the patient. For managing infected places, for example, if you visited and infected store later detected. The data, i.e. (Bluetooth id history) of the shopkeeper would also be analyzed, and if you have visited the store, your device mac address would be stored in the shopkeeper's device. The feature is possible due to the categorization algorithm of places mentioned in the 1st point of Data Analytics of the document.

**Feasible Extension of the idea to make it full proof**

**The idea could be further extended to make the public transport secure by installing very commercial Bluetooth receivers and broadcasters on transport like buses, metros or trains. This way, these devices would be storing the history of the people using the service.**

**Private data form cab & service companies like (Uber, Ola, RedBus, MakemyTrip) can also be used for more in-depth network analysis.**

**If the idea is further extended to an IoT device could be installed at stores and workplaces which keep track of their employees and customers. As a Bluetooth module is the main component used, industrial production would be feasible.**

**The implementation of these extensions may provide a deeper level of geostatistical social network analysis and help in better prediction.**

**Privacy of the users**

The data is encrypted on users end and only can be decrypted by the hash code stored on the user's phone. The code is only communicated through the official doctor's portal for marking the patient. When a patient is found positive, he is marked on the database. He is required to share his private key using a QR code generated on his mobile app. Then the encrypted data of the user is fetched and decrypted for analysis purpose. The encryption standards meet the industrial requirement and used by application like WhatsApp.

Our architecture supports this level of privacy and efficiency because we don't need to analyze the user's data unless found positive.

**Method for Statistical evaluation**

The data is used to plot and trace back the origin of infection and contain spread using an unsupervised and statistical probability-based calculation. The methods for data visualization included heatmaps, most places effected, spread rate calculation etc. We plan to use unsupervised Hidden Markov model for calculation of the probability of a person based on the network analysis of the infected with the rest of the population.

**Future scope of the app**

the app can be converted for more general tracking and health monitoring application. We have latently built fields to support other disease detection and analysis. However, further study is required for other diseases. Once implemented in full scale, we may achieve to identify and contain pandemics, epidemic, endemic and other diseases.

**Uniqueness from an existing product**

Arogya Setu

Arogya setu uses Bluetooth to detect people in contact and send geolocation data removing personal identifiers for creating heatmaps.

Arogya setu store data on the mobile app, and it is only shared when the person is found positive. This is done to prevent accuracy breach. Our app, on the other hand, uses encryption for keeping the data secure. Significant advantages of storing encrypted data on the server are:

1. It is proofed from general failure resulting in data loss. E.g. if someone uninstalls the app or phone/app gets some unseen internal error. The data is backed up on the cloud, and the privacy is also maintained
2. Local Storage of mobile is not used thus adding convenience at the user's end

Arogya setu only uses GPS data to create heat maps and quarantine places and location, which might be inaccurate due to the limitation of GPS and elevation of different places. E.g. if someone visits Mac Donalds which is on the 2nd floor of ambience mall. Arogya Setu cannot locate that the person visited Mc Donalds. The whole of ambience mall needs to be quarantined. However, using our architectures, we can locate exactly which shop/ store to quarantine.

From research, it has been found that coronavirus can remain on the surface and be suspended on the air for a specific time. Arogya setu is not equipped to detect people who have visited that infected store for the past days. Since the data is made private and we cannot access user's data; thus, complex geolocation backtracking is very difficult to perform in such cases. Even if data of users are made public backtracking of information would be very much complicated and computationally expensive.

In our solution, we can backtrack all people who might be at risk without trading privacy. We can have an option on the app for restaurants and stores to register. It stores data about all the customers who have visited using the same Bluetooth technology. However, in this case, all the data would be public. Social network analysis could be done on the patient's data to get the restaurants id (Bluetooth id of the restaurant/store). Thus, using the publicly available data of the store, we can backtrack all the people who might have visited the store after our patient.

There is not one app in the market which could detect possible infection in public transport. With our app not only we can detect people who might be at risk but also detect the bus/ car/ metro which needs to be sanitized. We are using advance human activity prediction (using machine learning) data which helps us to make more accurate prediction on the same.

coronatracker.in is an webpage which only plots data about the where abouts of the corona patients our app uses NLP on the map data to create a statistical model which mark all the possible infected location. This is done based on statistical analysis on the data of the patients received. Even if coronatracker.in plots data based on patients GPS location, data while transport will cause amiguity