Using Telecom Data for COVID-19 Contact Tracing

Samir Datt

skd@forensicsguru.com

ABSTRACT

2020 has been a year dominated by the spread of the COVID-19 virus throughout the world. Governments worldwide are experimenting with different ways of containing the virus spread. Many different methods have been tried with varying degrees of success. One of the thrust areas is the identification of contacts of COVID-19 positive patients. Due to the 14 to 15 day gestation period of the virus, a reverse method of contact tracing has to be implemented to determine a list of people who may have had contact with this infected person and could probably be infected. These potentially infected people would need to be quarantined to further contain the virus spread. This paper discusses a number of scenarios and the way telecom log data can be used to identify and trace the potential contacts of COVID-19 positive patients.

INTRODUCTION

In the month of March 2020, there was a meeting of the Tablighi Jamaat (a Muslim Missionary Movement) in the Nizamuddin area of New Delhi. This was the largest gathering of its kind. There were visitors from all over the world, including countries that had already been infected by the rapidly spreading COVID-19 virus. As the meeting came to an end, most of the infected including over 1000 foreigners who had attended the Markaz, dispersed to different parts of India carrying the virus with them. A large number of them attempted willfully to avoid quarantine and were identified as actively spreading the virus. At one stage over 25,000 Jamaat members and their contacts had been quarantined across nearly 15 states.[1]

The Government of India, launched the Aarogya Setu app with the intention of contact tracing. However, the app suffered from some inherent drawbacks. It required Bluetooth to be turned on at all times. This acted as a deterrent to many because Bluetooth was observed to be a major drain on battery life. The second

1. https://www.aljazeera.com/news/2020/04/tablighi-jamaat-event-india-worst-coronavirus-vector-200407052957511.html

factor that affected use of this app was its optional nature. A lot of people did not install the app and hence the Aarogya Setu app was not as effective as was initially envisaged. The Tablighi Jamaat members made it a point to avoid installing apps of this sort and thus made it extremely difficult to trace them and contain the spread of the virus.

It was seen that a number of international governments were employing Telecom data analysis methods for contact tracing of COVID-19 Infected persons. However, details of such methods were not easily available though they were hinted at in a number of news articles. [2]

To handle this situation and enable contact tracing of people who did not want to be traced, the author of this paper has researched a number of methods by which possibly infected contacts of Covid-19 infected patients can be traced using Call Detail Records (CDR), Tower Dump Records (TDR) and Internet Protcol Detail Records (IPDR) data obtained from telecom logs requested from telecom service providers by Law Enforcement agencies via due legal process.

As of the 16th of August 2020, there were over 2.6 million COVID-19 confirmed positive cases in India, with nearly 6,80,000 currently active. The effect of the Corona virus has been widespread and extremely detrimental to the health, sentiment and economy of the country.

---

2. https://www.business-standard.com/article/current-affairs/in-pics-how-nizamuddin-became-the-biggest-coronavirus-hotspot-in-india-120040201078_1.html

.https://economictimes.indiatimes.com/news/politics-and-nation/maharashtra-deploys-drones-uses-cdr-boosts-patrolling-to-contain-covid-19/articleshow/74999989.cms?from=mdr

https://www.sciencemag.org/news/2020/03/cellphone-tracking-could-help-stem-spread-coronavirus-privacy-price#

IMPORTANCE OF THE CONCEPT

The District Administrations under the leadership of the District Magistrate [DM] in close conjunction with the Police has been tasked with containing this pandemic. Every state, in fact every district is attempting to tackle the problem in their own way with varying degrees of success. The urgency and magnitude of the pandemic requires concerted efforts from all quarters to mitigate the effects of this devastating pandemic.

In this context, this paper is a very important resource to further the fight against the virus by using technical methods in an effective way for tracking, tracing and containing contacts of COVID-19 positively affected persons. If these methods are deployed across every district in the country, we can expect a much better containment of the virus spread and a consequent reduction in the virus infection rates. The best results can be obtained by sharing the methods contained in this paper with the widest possible audience in the least possible times, especially the police establishments involved in contact tracing as well as the District Administrations involved in coordinating these efforts.

METHODS

Due to the complexity of the problem and the different approaches that are required to be taken, six different scenarios are discussed and the steps that need to be taken for the analysis of telecom data for the purposes of contact tracing are presented.

For the purposes of analysing the data the following tools were used:

- Cell Site Analyser CSA™ - for identifying active cell tower coverage in an area
- Call Data Analysis & Management System - CDAMS™ - CDR, Tower & IPDR Analysis
- Analyst Notebook - i2 ANB ™ - Link Analysis

SCENARIO 1 -

"Identified COVID positive person, not cooperative, un-willing to disclose contacts & does not have Aarogya Setu (Contact tracing app) on phone"

In such a scenario the identified COVID infected person (such as members of the Tablighi Jamaat) is withholding information and attempting to thwart law enforcement efforts to identify his contacts and as such an alternative method of identifying his contacts has to be employed.

The following steps are employed in this scenario -

1. Obtain CDR of the COVID-19 positive infected person.

   Once a COVID positive person has been identified, the concerned Police department identifies his/her mobile phone number and then puts in a request to the relevant service provider for his CDR (Call Detail Records) of the last 15 days.

   A CDR usually contains a detailed log of all the communications/interactions between the specified number (of the COVID positive person) and other parties. These interactions could be in the form of a call, a text message or just internet related activity.

   A CDR usually consists of data structured in a number of columns. These usually are -

   Target number - This is the number that belongs to the COVID+ ve person

   Other Party number - This number is the number of the correspondent with the target number. This could belong to friends or family or just an acquaintance. Sometimes these are machine/computer generated and are identified by unique codes called "Short Codes"

   Date of call - The date when the call was made - every service provider provides dates in a multitude of formats.

   Time of call (usually in 24 hour format but service providers do have different formats including AM/PM and others)

Duration (usually in seconds, but some service providers provide this in different formats such as hh:mm:ss)

Call Type (This is usually in abbreviated code and also designates the direction of the call) - eg. SMT - Short Message Terminating (denotes an incoming SMS)

Starting Cell Id - (Cell Tower Id at the start of the call - basically the general location of the person while initiating the call)

Terminating Cell Id - (Cell Tower ID at the end of the call - just like the above - this denotes the general location of the person at the time of the end of the call)

IMEI - The International Mobile Equipment Identity - a serial number that is unique to every phone instrument - the first few digits when deciphered correctly help in identification of the make and model of the phone device. Hence, once the make and model is known it is possible to identify whether the phone instrument is a dual sim device or not. In dual SIM systems each of the SIM slots on the phone is allocated a unique IMEI number. Thus once one IMEI is known, the other can be determined from the mobile device seller and then the IMEI can be communicated to the Telecom Service Providers to get the associated CDR.

IMSI - The International Mobile Subscriber Identity - This is the unique serial number that is allocated to the SIM by the Mobile Service providers. In dual Sim systems each of the SIMs is allocated an individual IMSI.

To better understand locational information associated with a CDR, a chart known as a "Cell Id chart" is also requested from the service provider. This chart lists out the address and latitude and longitude associated with every Cell Id number in the region. Thus, the exact geographic location of every cell tower is known. Further the azimuth helps identify the direction (the arc of coverage) of the cell tower in question. The data from the Cell Id Chart is merged with the CDR to get relevant locational data of the COVID infected person at the relevant time. This data is easily provided by

service providers to Law Enforcement on their official request as per mandated procedures and appropriate authorisations.

2. Identify top contacts of this person on the basis of call frequency and call duration.

Once the Targets call detail record is obtained, it is seen that each call/ interaction is represented by a row of data. A detailed analysis of the rows is done and a list of numbers that have multiple interactions with the target are identified. These are then ranked on the basis of -

. The number of times they have interacted with each other in the specified period.

b. The sum of the durations of all the interactions between the two parties.

The top 5 "other parties" for both the use cases are identified. This set of people are now potential contacts for the COVID infected suspect. To narrow the list further down we need to look at their locational proximity to the COVID positive person during the previous 15 days. Hence we need to identify the specific locations visited by the COVID infected person as well as the locations of the "other parties".

3. Identify top locations of these persons on the basis of day & night location analysis.

Similarly, a count of the most visited locations (based on the Cell Ids) during the night (8 PM to 8 AM) and during the day (8 AM to 8 PM) is made. These locations are ranked on the basis of highest visit frequency.

It is assumed that since the COVID positive person (target) has spent considerable time in these locations with a high frequency of visits, the probability of spread of the COVID virus from contacts at these locations is higher. Thus, in the next step we attempt to identify those of his top contacts shortlisted earlier who are in the vicinity of these top visited locations by the target.

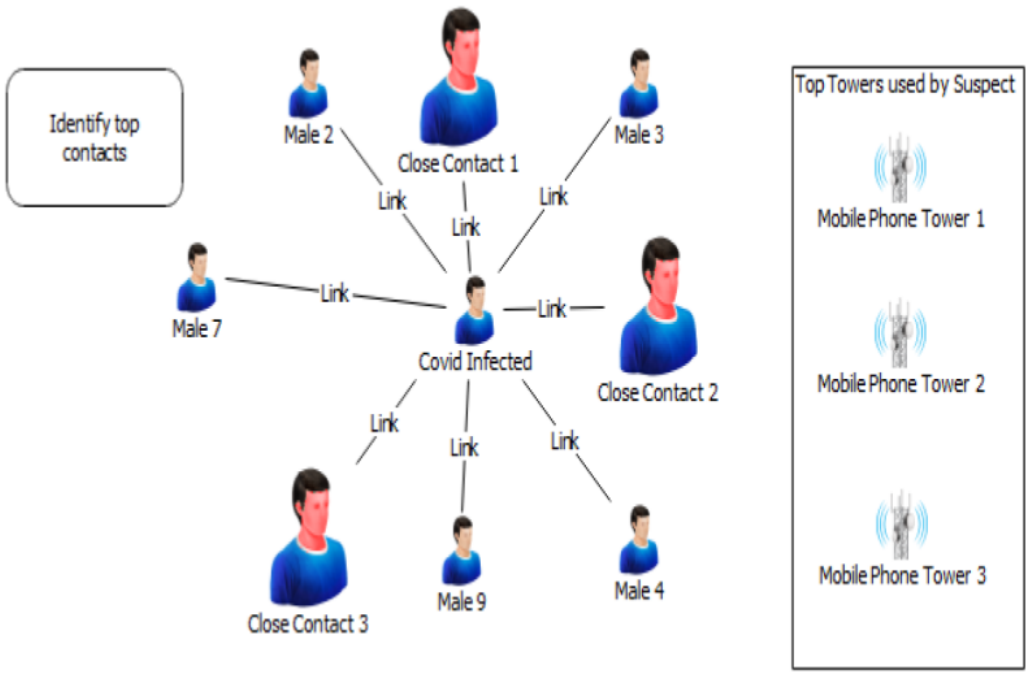

Figure 1: Pictorial representation of calls of a Covid infected person as well as his top locations from cell towers. Close contacts have been intentionally made larger with a red coloured complexion.

4. Obtain CDRs of top contacts of the COVID-19 positive person.

Law Enforcement requests Call Detail Records of the earlier identified top contacts of the target during the same 15 day infectious period. In the next step a location analysis of these contacts is done.

5. Identify the top contacts location within 2-5 Kms of COVID-19 positive infected person's location.

The CDRs of each of these contacts is plotted separately on a map in conjunction with the locations of our target during the same period and their relative geographical location is determined. This data is plotted on maps using the latitude and longitude information associated with the cell tower ids obtained in the CDR. If any contacts are found within a radial distance of up to 5 kilometers, the contact is approached by law enforcement over the telephone and some basic questions are

asked related to his/her association with the target and their relative movements during that period as shown in the figure 2 below.

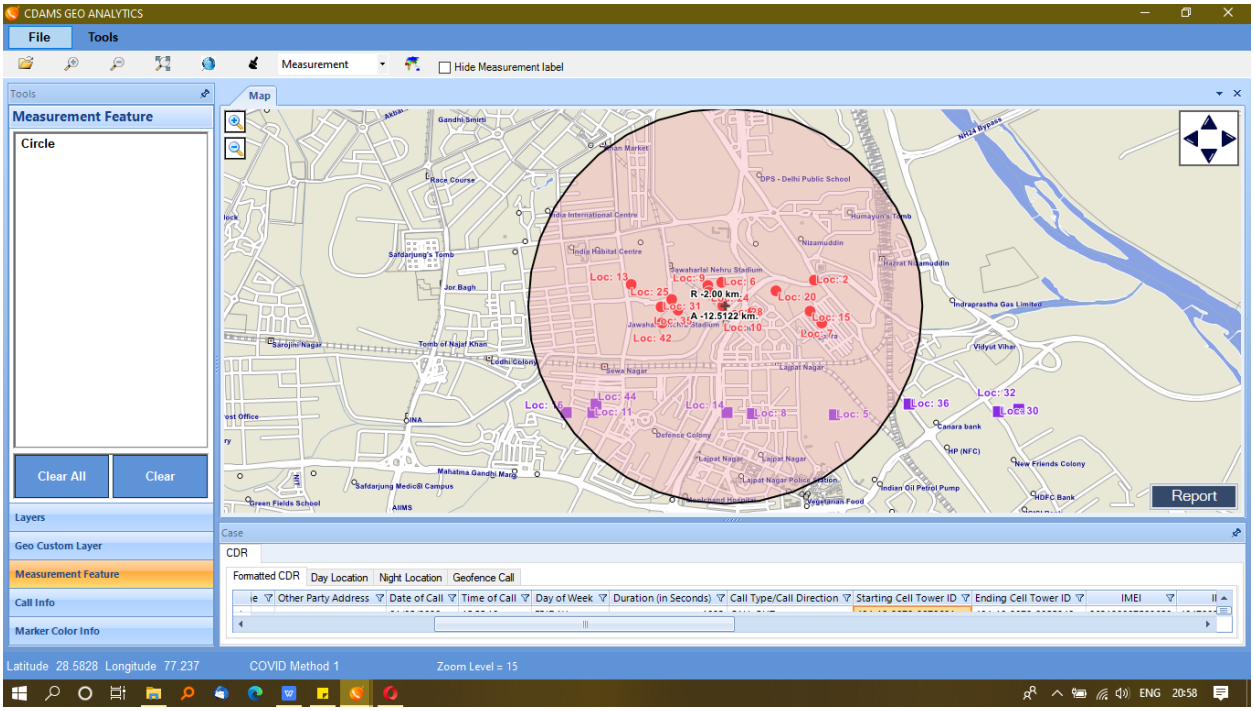

Figure 2: The Geo Analytics module of CDAMS showing cell tower proximity of contacts.

6. If found, person to be questioned & quarantined.

If the contact is cooperative and currently does not have symptoms, the contact is tested and advised to be home quarantined due to their contact with the target as a precautionary measure.

If the contact is found to have symptoms and tests positive for the virus, appropriate measures for containment and cure are put in place and a next level contact tracing exercise for this person is also initiated.[3]

3. https://www.zdnet.com/article/coronavirus-they-want-to-use-your-location-data-to-fight-pandemic-thats-a-big-privacy-issue/

https://spectrum.ieee.org/tech-talk/telecom/security/how-coronavirus-pandemic-europe-collecting-phone-data

https://www.businessinsider.in/defense/news/10-countries-are-now-tracking-phone-data-as-the-coronavirus-pandemic-heralds-a-massive-increase-in-surveillance/articleshow/74744866.cms

Covid 19- https://techcrunch.com/2020/03/27/telco-metadata-grab-is-for-modelling-covid-19-spread-not-tracking-citizens-says-ec/

https://www.theverge.com/2020/3/23/21190700/eu-mobile-carriers-customer-data-coronavirus-south-korea-taiwan-privacy

In some situations, the contacts of the target are also not cooperative and do not show any inclination to cooperate with the authorities. In these kinds of cases, the address of the possibly infected contact is determined from the CAF (Customer Application Form - submitted to the service provider at the time of purchasing the mobile number). In situations where the address from the CAF is found to be false or the contact is no longer (intentionally or unintentionally) at that address, an exercise to locate the contact has to be initiated. This is usually done by examining the CDR further to determine his usual night halt and day halt locations and also using short code analysis to identify food and online delivery services which have been in touch with the missing contact. In depth analysis of this data allows the identification of the exact address being used by these possibly infected individuals. Additionally, trails left by the usage of electronic money transfers, ATM withdrawals, CCTV footage etc. can be used to trace such contacts that are "on the run". [4]

SCENARIO 2 -

"Having Identified a group of COVID+ suspects (eg. could belong to a gang), we need to identify possible contacts affected by them"

This is a fairly straight forward exercise. The following steps are followed -

1. CDR's of all the COVID Infected persons (members of religious groups/ gangs who refused to be tested/located) are collected.

   In this case law enforcement requests CDRs of the last 15 days of the persons identified as COVID infected. As all the CDRs are not likely to be from the same service providers the CDRs need to be normalised (brought into the same columnar structure) and then analysed.
2. Common other parties of these infected persons are identified

---

4. Tracing the virus using Applications- https://www.forbes.com/sites/bernardmarr/2020/04/09/the-vital-role-of-big-data-in-the-fight-against-coronavirus/#69cb45f33806

The other parties of all the COVID infected targets are examined and it is determined if any of the other parties are common to any/all of the CDR's. Parties common to two or more CDR's denote that the common party is known to both (or more) infected persons and as such the probability of infection of this person can be quite high. (See Figure 3)

Further tests such as proximity analysis can be done to find if the possibility of infection goes up or down.

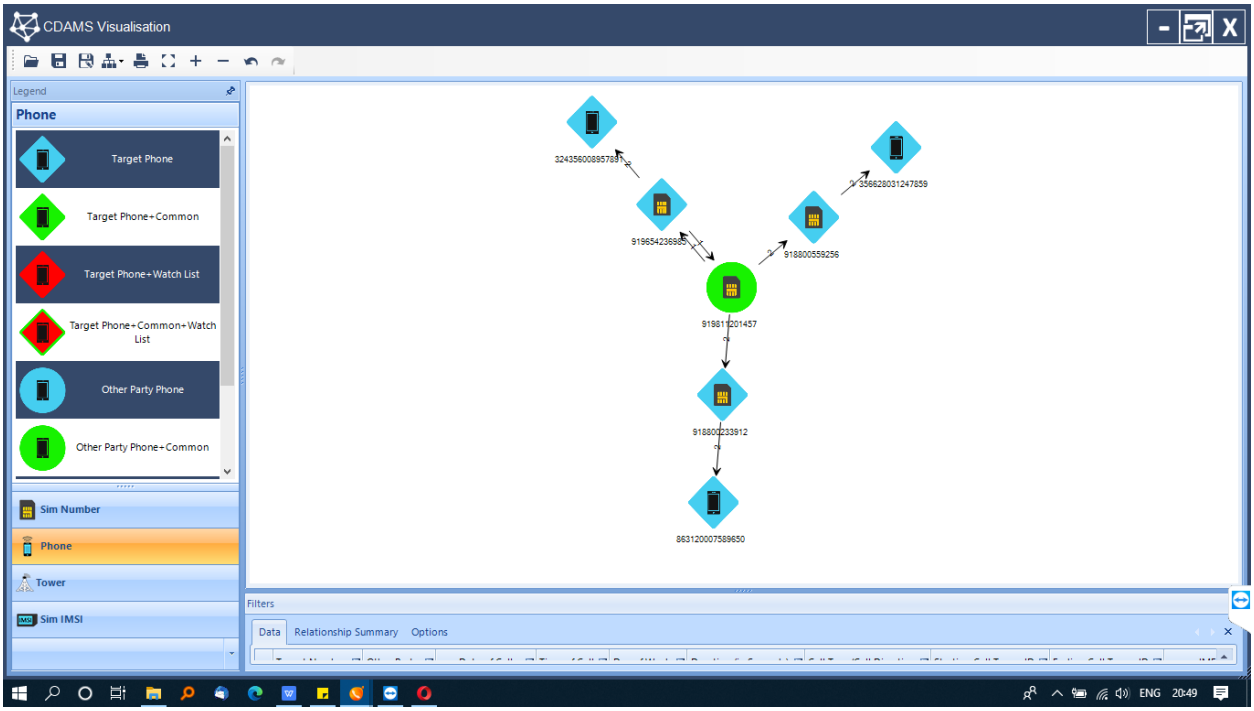

Figure 3: Depicts common contact (number) between three COVID positive persons

3. These common persons are tested and quarantined.

Similar location procedures as specified in Scenario 1 are followed.

In Figure 3 above, the end points represent COVID positive persons and the green icon represents a hitherto unknown contact common to all the infected persons and possibly infected as well.

SCENARIO 3 -

“Needing to identify possible out state (travellers) COVID positive suspects crossing a state border and entering the state.”

When knowledge of the virulent nature of COVID became widely known, there was a major exodus of migrant labour from the cities to their home states. Unfortunately, a lot of these travellers became virus carriers and ended up becoming a major containment nightmare for Law Enforcement Officers of their home states. A number of these evaded detection and crossed state borders on foot, on bicycles, in the back of trucks etc. The exodus has now reversed with the migrant labour heading back to the cities in search of livelihood.

The proposed method leverages the fact that the moment a cell phone carrying person moves out of one state and into another, the Mobile Service Provider (MSP) sends the phone a sms welcoming the person to the state and informing them of entering the new zone [see figure below]. From the MSP point of view this is done from a billing perspective, however this can be quite useful from a law enforcement perspective as well. These text messages are sent from MSP owned “short codes” and are computer generated messages. These “short codes” are identified and documented and a list of people entering the state can be determined by getting a list of numbers that have entered the state by looking for these short codes in tower dumps and CDRs of suspected individuals. See figures 4 and 5 below

Figure 4: SMS received when entering the state of Haryana

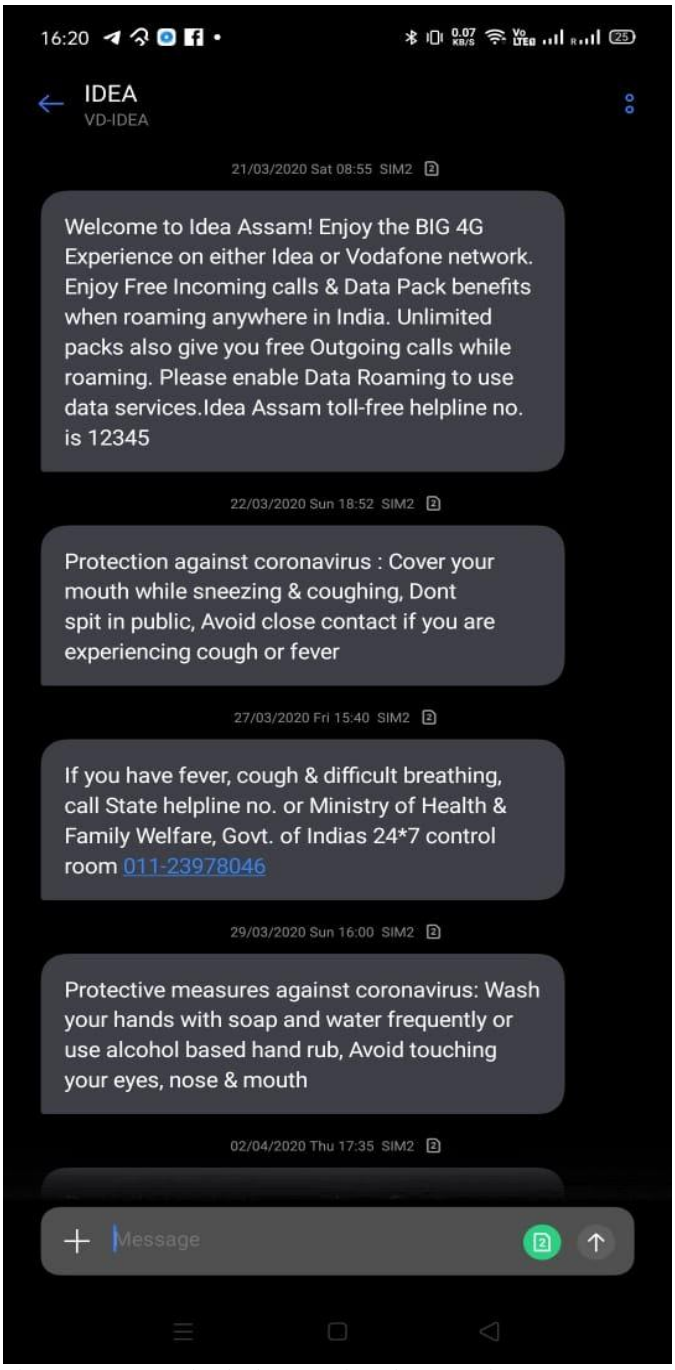

Figure 5: SMS received when entering the state of Assam

The following steps can be followed -

1. Obtain CDR of COVID positive person or TDR of border Cell Towers from MSP

   Once a COVID positive person is identified, his CDR is requested from the concerned service provider.

2. Identify system generated SMS (from their “short code”) showing entry into the state. Note their date and time

The targets’ (COVID positive person) CDR is scanned for the presence of service provider short code messages as shown in the pictures above. These usually contain a welcome message, welcoming the subscriber to the state. The date and time of these messages are noted and are used as a frame of reference to identify others who may have crossed the border at the same time.

3. At the same date and time identify others who also have crossed the state border. These may have travelled together by bus or car or other modes of transport.

A list of numbers that have been sent a welcome message by the service providers at that border are requested for and are filtered on the basis of the above determined time and date.

The responding number owners are contacted and questioned as to their whereabouts at that time as well as their acquaintance with the infected person.

4. Test/Quarantine such identified others.

Those persons that fit the above filter criteria are tested and quarantined as deemed fit by district health and LE personnel.

SCENARIO 4 -

“Alerting Possible contacts about the possibility of having been in the vicinity of a COVID positive person”

The nature of the Corona virus makes it quite difficult to detect and in a number of cases the infected person may be asymptomatic. When a person is identified as COVID positive, it becomes imperative to identify the person's movements over the previous 15 days to identify the people he/she may have come in contact

with. This is depicted by the red dots in Figure 6 While this is comparatively easier for known people, it becomes extremely difficult in situations where random movement has been made over the previous “risk” period.

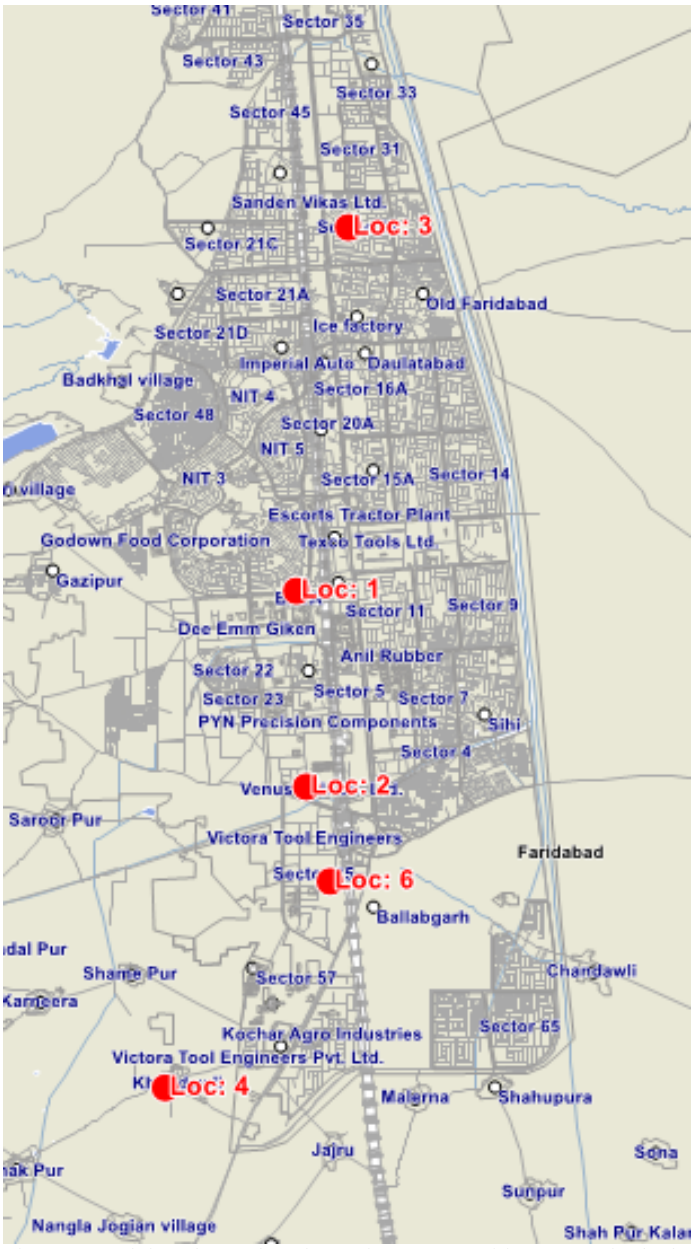

Figure 6: Red dots denote locations of COVID positive person in the last 15 days

The proposed investigation method is as follows -

1. Obtain CDR of the previous 15 days of the infected person.

A request will need to be made to the MSP for the CDRs of the infected person. The contents of the CDR would need to be analysed for locations relative to dates and times.

2. Identify this person's top locations on the basis of day and night location analysis

A detailed location analysis will show the major locations the suspect has been to during the spread period. Each of these locations would need to be documented in conjunction to the corresponding dates and times. See Figure 7

3. Identify all cell towers of other service providers in the vicinity of the identified locations above.

Based on the data obtained above, cell towers in the vicinity belonging to other MSP's are also noted. This comprehensive list of cell towers is sent to their respective MSP's to obtain the corresponding tower dumps for the concerned period.

4. Obtain tower dumps from MSP's for these identified cell towers

The MSP's are requested for the data corresponding to cellular usage in these towers during the specified period. This is usually provided on a whole day basis.

5. Build a filter around the date and time of movement of the infected person and filter out people who were not present in the vicinity at the specified date and time.

Based on the known date and time of the infected person's movements, filters are applied to identify people who were in the vicinity of this person at the time. A list of such "at risk" persons is compiled.

6. An automated "appropriately worded" message requesting self quarantine is sent to the "at risk" people to reduce the risk of onward infection and community spread.

This can be done in a manner to recommend home quarantine, which would further restrict the spread of the virus.

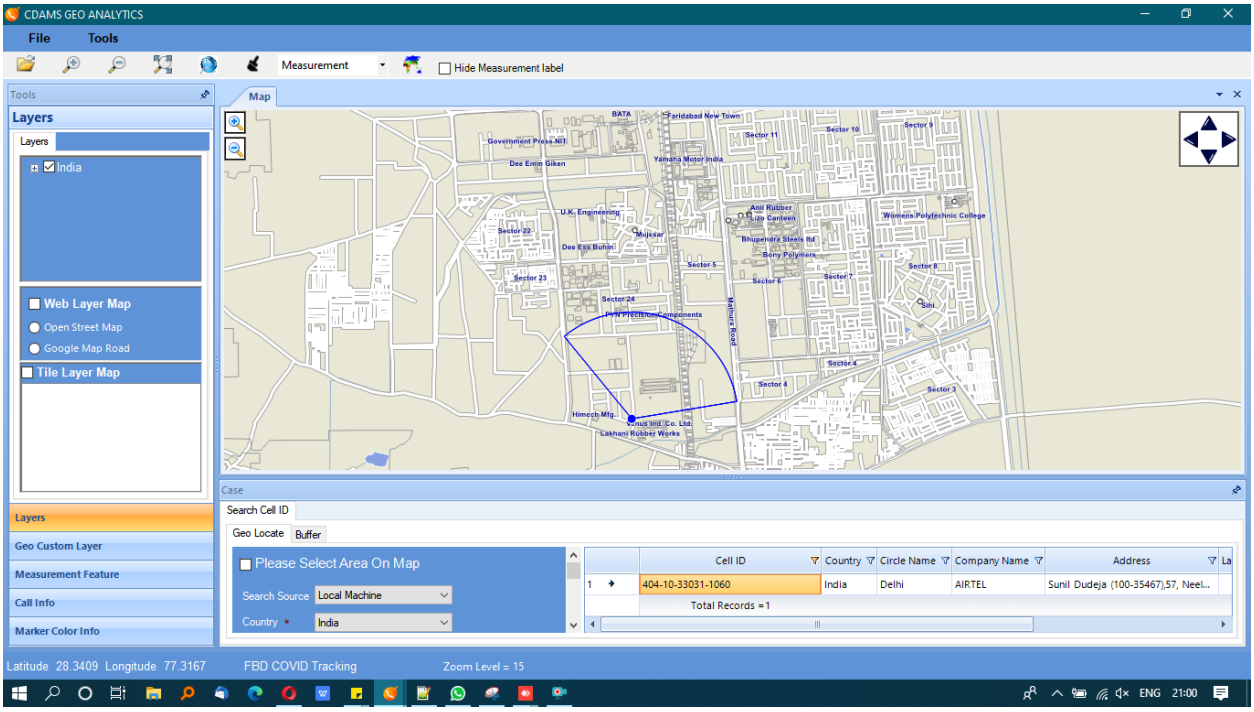

Figure 7: CDAMS Geo Analytics module showing the area covered by a single tower

SCENARIO 5 -

“Quarantine Zone (Red Zone) Monitoring”

With the increasing number of cases of COVID-19, certain areas of cities are designated as “Quarantine Zones”. These zones are basically NO GO zones - with no movement allowed to and fro from them for a predetermined period of time. Since these areas are high density infected zones it is critical to monitor movement to and from these zones. It is also essential to identify people and track people who come and go from this zone.

The following steps can be followed to monitor Quarantine zones.

1. Identify Cell towers that cover the quarantined area

A Cell Site Analyser (CSA™) device is used to survey the zone to be quarantined. This device identifies and lists the cell tower id of all the cell sites (irrespective of service provider) that cover the "Quarantine Zone". See Figure 8. Requests for tower data are sent to the MSPs of the respective cell towers pertaining to a 15 day period prior to the beginning of the Red Zone Lockdown.

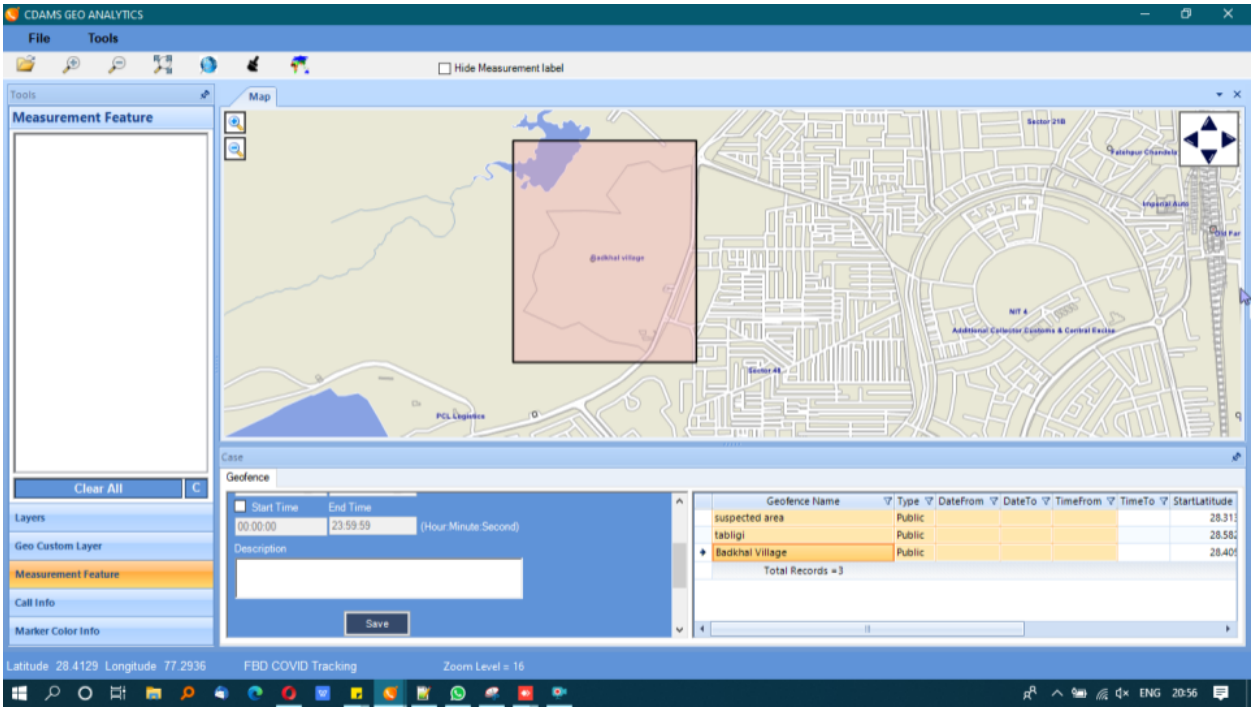

Figure 8: CDAMS Geo Analytics GEO Fencing module covering the Quarantine Zone

2. Using Tower dumps to identify the regular residents of the area

These tower dumps are analysed over the specified time period and an indicative list of telephone numbers that make calls from the area during night time on a regular basis is compiled. These people (telephone numbers) would comprise the regular residents of the area in most cases. This set is used as our reference set for further analysis. This is a big data problem and usually requires special tools to do the heavy lifting. In this particular case CDAMS™ Ultimate Edition with Microsft SQL Server in the backend was used.

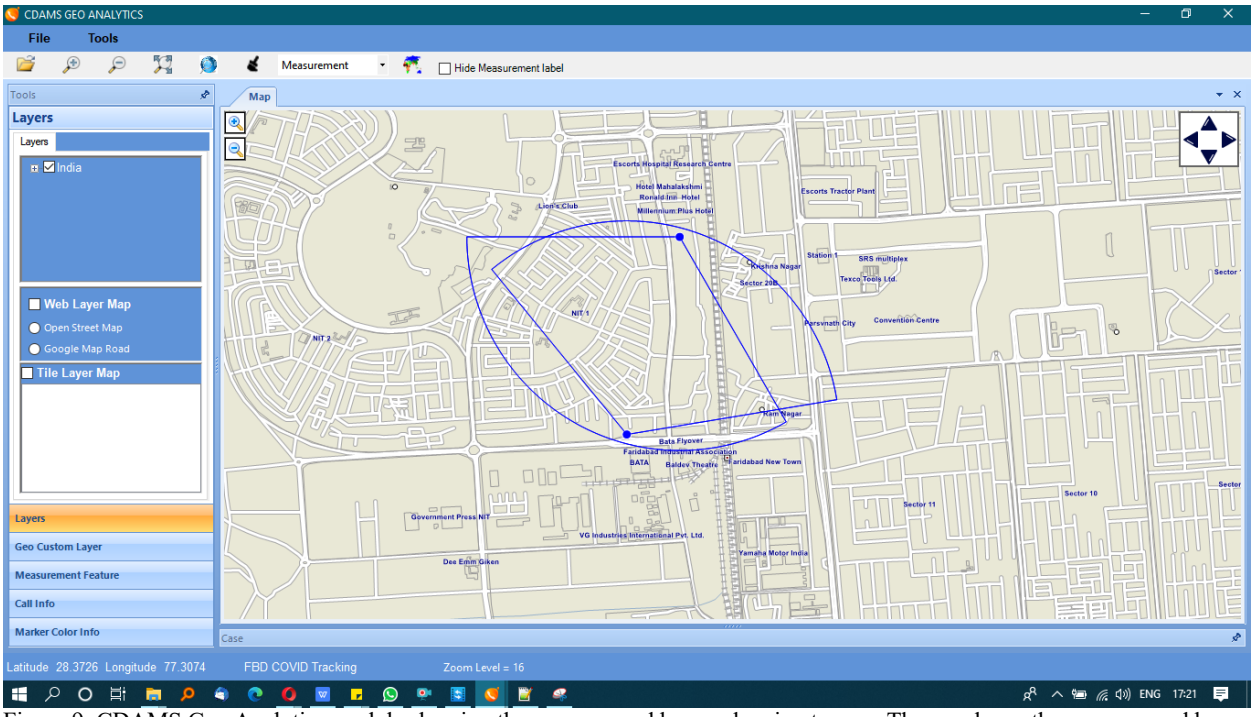

Figure 9: CDAMS Geo Analytics module showing the area covered by overlapping towers. The arc shows the area covered by each tower as calculated from their azimuth information provided by the mobile service provider.

3. Identify one time Leavers - People exiting the quarantine zone

Another set of data is requested from service providers post the implementation of the quarantine. Using the above reference data set as a control, analytics is run to identify residents who leave the area covertly once the quarantine has been initiated. These are likely to be infection vectors and would need to be traced/ monitored. Their individual CDRs would need to be requested from the service providers and contact tracing scenarios as mentioned above would need to be implemented.

4. Identify one time Joiners - People Entering the quarantine zone

It is fairly common that outstation people upon hearing of a lockdown in their areas tend to return to their homes with the objective of being with their families etc. Not only can this bring in additional infected persons into the area, it may also increase the spread of the virus to those who are hitherto un-infected, who have opted to come into the quarantine zone. Hence an analysis of the second set of tower dumps relative to the previously identified control group of residents would

show which numbers have been added post the lock down period to the residents pool. These "Joiners" or "new residents" can be home / self quarantined to help keep things in control.

5. Identify Circulators - People who leave during the day and return at night or vice versa

In any kind of a lockdown there will always be people who would like to leave the controlled area for their work of the day then return back to it in the evening. A number of these people will try and escape the quarantine and attempt to sneak back in at night. A number of these will be successful. However each of these "Circulators" will carry the risk of spreading the infection thereby defeating the very purpose of the quarantine.

Identifying a "Circulator" is more complicated than the other options. In such a case an area larger than the Quarantine zone will need to be geofenced and data from the cell towers covering the areas from the end of the "red zone" to the outer "super set" geofence will need to be requested from the service providers. This data would need to be checked for any of the residents, to enable us to identify the people leaving the quarantined area and then returning back to it later. The presence of these "circulators" in these out of quarantine area cell towers will help in zeroing in on them.

Once identified standard containment procedures can be applied.

SCENARIO 6 -

"Identify Infectors presence in densely populated places, such as malls and shopping complexes"

Watchlist Alerting system

A list of all known Corona positive active patients/carriers and their phone numbers is made

This list is made as a watchlist

In densely populated areas such as markets/malls etc - tower dumps are taken and the Watchlist alert system throws up Alerts in case the COVID carriers are seen to be roaming around in those areas at that time rather than staying home quarantined.

Procedure to be followed -

1. Take a regularly updated list of COVID-19 positive infected persons and their contact numbers.

   District administrations have constantly updated records of infected persons in their areas. From the time a person is diagnosed as COVID positive till the time the person is retested to be COVID negative, a record of the infected person is available with the District Administration. Such identified persons are required to stay either quarantined at home or in hospitals. Their mobile numbers are fed into a watchlist.

2. Identify the COVID-19 positive infected persons in highly populated areas like- markets, Shopping malls, Food Courts.

   Densely populated areas such as markets, malls etc are a potential hot spot in terms of virus spread. From that perspective it is important to prevent the free movement of COVID infected personnel in these areas or zones. A cell site survey is done for these areas and the cell towers covering these areas are identified. Tower dumps are requested and the watchlist compiled earlier is run against these dumps to identify “infectors” wandering about in these areas on an unauthorised basis.

3. If found, the person to be questioned & nearby persons should be notified.

   If the watchlist triggers a presence alert, the government and police departments can swing into action and proceed to detain the infected person and identify further movement as well as notify people who may have had exposure to the virus.

FUTURE DIRECTIONS

Due to the nature of the problem, its complexity and the rapid way in which the infection spreads, this paper perforce has to be a "work in progress". The relationship/ collaboration between the district authorities/ health, law enforcement and telecom service providers will need to be strengthened. Certain modifications may be required in the legal frameworks pertaining to the collection of telecom data from service providers keeping privacy issues in mind. [5]

Ongoing research relating to new methods of identifying both the infected and their contacts with even higher degrees of accuracy is required. Closer cooperation of the authors with both Law Enforcement agencies as well as Mobile Service Providers is necessary to enable realistic research and determination of additional methods that could assist in the fight against this pandemic.[6]

---

5. https://economictimes.indiatimes.com/industry/telecom/telecom-news/telcos-dot-in-sync-with-states-to-track-patients/articleshow/75395651.cms

Covid 19 & App 'electronic fence'- https://www.reuters.com/article/us-health-coronavirus-telecoms-eu/vodafone-deutsche-telekom-6-other-telcos-to-help-eu-track-virus-idUSKBN21C36G

[6] ACKNOWLEDGEMENTS

This paper owes its existence to the constant encouragement of Prof. Arvind Verma who has been the inspiration behind it. Prof Suresh Lodha has inspired the visualisations. I am also extremely grateful to a number of police officers who have been very helpful in sharing ideas and methods they have used in tracing contacts. I would also like to acknowledge the rock solid support from the talented team at ForensicsGuru.com who helped me work out different scenarios for contact tracing.