A Privacy Preserved and Cost Efficient Control Scheme for Coronavirus Outbreak Using Call Data Record and Contact Tracing

Shibli Nisar, Syed Muhammad Ali Zuhaib, Abasin Ulasyar and Muhammad Tariq

Abstract—Coronavirus or COVID-19, which has been declared pandemic by the World Health Organization, has incurred huge losses to the lives of people throughout the world. Although, the scientists, researchers and doctors are working round the clock to develop a vaccine for COVID-19, it may take a year or two to make a safe and effective vaccine available for the world. In current circumstances, a solution must be developed to control or stop the spread of the virus. For this purpose, a novel technique based on call data record analysis (CDRA) and contact tracing is proposed that can effectively control the coronavirus outbreak. A positive coronavirus patient can be traced through CDRA and contact tracing. The technique can track the path traversed by the patient and collect the cell numbers of all those people who have met with the patient. Keeping in tact the privacy of this group of people, who are contacted through their cell numbers so that they can isolate themselves till the result of their coronavirus test arrives. If a test result of a person comes positive among the group, then he/she must be isolated and same CDRA and contact tracing procedures are adopted for that person. A COVID-19 patient is geo tagged and alerts are sent if any violation of isolation is done by the patient. Moreover, the general public is informed in advance to avoid the path followed by the patients. This cost effective mechanism is not only capable to control the coronavirus outbreak but also helps in isolating the patient in his/her house.

Index Terms—CDRA, Coronavirus, Control, COVID-19, Cellular Forensics, Hidden Pattern, Privacy Preservation

1 INTRODUCTION

Coronavirus disease (COVID-19) is an infectious disease, which is caused by a newly discovered coronavirus [1]. So far, the disease has spread like a wildfire and almost reached to the entire world by affecting millions of people [2]. The most scary side of the disease is that, it is a novel virus accompanied with unknown myths and realities [3]. Covid-19 belongs to a family of ribonucleic acid (RNA), which is basically a virus family named as coronaviridae [4]. To make the curve of coronavirus outbreak flatten, the governments have imposed lockdown in their countries, enforced border shutdowns and travel restrictions. Consequently, many social and economic sectors are affected such as petroleum and oil industry [5], agriculture [6], manufacturing industry [7], education [8], finance industry [9], healthcare and pharmaceutical industry [10], tourism industry [11], real estate and housing sector [12], sport sector [13], information technology and media industry [14]. Moreover, social distancing and lockdown measures to control the spread of virus have caused severe impacts on social life in terms of domestic violence, that includes emotional, physical and sexual abuse [15].

With the emergence of coronavirus, researchers from all over the world are actively engaged to find the drug or vaccine and discover different ways to control its spread. Since, there is no vaccine or drug available to cure this deadly disease, therefore, the best solution is to control the wide spread of the disease by social distancing and isolation [16]. Currently, the main concern is to minimize the coronavirus outbreak. In order to control the spread of the virus, this research work proposes a novel solution that can track, trace and terminate coronavirus disease by using call data record analysis (CDRA), which is commonly known as call detailed record. A call data record is a log file maintained by the telephone or mobile operators. The log file contains data such as, duration of calls made or received at a particular time, source and destination numbers, cell type, international mobile equipment identity (IMEI), cell site detail, and latitude and longitude of base transceiver station (BTS) [17].

Almost all law enforcement agencies (LEA) use call data record (CDR) for tracking, tracing and investigation of criminals [18]. This is one of the most widely used techniques to cope with the criminals and have shown very promising results [19]. LEAs use CDR analysis to track and trace the entire criminal network and find out the hidden connections between criminals. This paper proposes a technique that assists the governments to track and trace COVID-19 patients with the help of CDRA in the same manner that is usually done for criminals. The concern department make it sure to keep the privacy of a patent in tact. The technique not only identifies the suspected cases through call data record of a patient but also it tracks the path followed by the COVID-19 patient.
One of the main limitations of the CDRA is when a suspect goes into the crowd without making a phone call. In order to overcome this limitation contact tracing based on exposure notification system is used in parallel to track the path traversed made by the patient without call history. The exposure notifications system (ENS), originally known as the privacy-preserving contact tracing project, is a framework and specification developed by Apple and Google to facilitate digital contact tracing during the COVID-19 pandemic. Hence the proposed model is capable to track and trace COVID-19 patients with and without call history as shown in Figure 1.

Rest of the paper is organized as follow: Section 2 presents the CDMA framework. Section 3 discusses contact tracing framework. Section 4 presents the conclusive remarks, potential outcomes of the proposed model, and some future directions.

2 CDRA FRAMEWORK

In order to curb the spread of COVID-19, a model is required that can immediately locate the coronavirus patients by using the geo location. This type of location can be found with the help of tower signals and call history of a patient. If we track, control and isolate, the spread can be controlled in a few weeks. Otherwise, if we ignore instruction at the start, it will be an immense damage. This can be easily done through the CDRA. Through CDRA, one can easily track the location and call history of a patient and can reach all possible suspects who met the patient. By doing so, one can easily stop or minimize the coronavirus outbreak. In addition, the potential possible carrier should be geo tagged and tracked by sending alerts if he or she violates quarantine or self-isolation. This will help the people to isolate themselves and not becoming the source of coronavirus spread. Detailed flow diagram of the proposed model is shown in Figure 2.

A CDR provides metadata. Metadata is a set of data that describes and gives information about other data. Such as how a specific phone number and/or user is utilizing the phone system. This metadata includes different information about the call but more relevant information it contained is summarized in Table 1.

CDRs can also include short messaging services (SMS) messaging metadata and any other official communications transmission. However, the contents of the messages/calls are not revealed through the CDR, thus keeping the privacy in tact. The call detail record simply shows that the calls or messages took place, and measures basic call properties. Initially, one has to perform CDR analysis on all those people who had recently traveled from abroad and could be source for spreading coronavirus in other people or those whose results are declared positive. If the patient is diagnosed with coronavirus, the hospital authorities will inform the government authority to perform CDR analysis on the patient’s active sim with the consent of patient. After CDRA of a patient, the concern government team will contact people with whom the patient might be interacted (these people are known to be suspected people). Ambulance service will go to suspected patients’ home and will get sample for coronavirus. During the process, suspected people will be

---

TABLE 1

| Call Data Record                                      |
|------------------------------------------------------|
| When the call took place (date and time)             |
| How long the call lasted (in minutes)                |
| Who called whom (source and destination phone numbers)|
| What kind of call was made (inbound, outbound, toll-free)|
| International Mobile Equipment Identity (IMEI) detail|
| How much the call cost (based on a per minute rate)  |
| Information about Base Transceiver Station (Location)|
| Longitude and latitude of BTS when call is made/received|
guided to stay at their homes or remain isolated until final result of the test is revealed. If the test comes positive, the patient will be sent to a hospital. Similar procedure will be repeated for all those people to whom the patient has been interacted and their corona test was positive. The potential outcomes of the proposed model are listed below:

- tracking of COVID-19 patients,
- monitoring of isolated patients,
- tracking of suspected ones,
- inform the mass about the safest path to use.

2.1 Case Study 1

In order to integrate the information and interpret the outcome of the proposed model, one complete case study based on fiction is discussed in this section. If the subject’s corona test comes positive then the very first step is to quarantine the subject and asked the mobile operator for his/her CDR. CDRA starts just after getting the CDR from the mobile operator. After receiving the CDR, it starts in the format given in Table 2. The table contains the most relevant attributes of the CDR.

CDR of the subject’s contained all the relevant information through which one can easily track and trace the patient through his/her cellular contacts along with path followed by that subject. From CDR, we can get the information about the calls made and received by the patient along with date and time. CDR contains the duration of all calls made and received. CDR includes information about the BTS with which the subject is connected. CDR also holds the latitude and longitude of BTS when a call is being made or received. Before starting the CDR analysis the very first step is to remove the past irrelevant data. This can be easily done by deleting all past entries and keeping the data of few couple of weeks. Doing so, one can easily limit the CDR data in the desired dates. This will help to analyze the small amount of data efficiently and accurately. In order to clearly visualize the data and extract hidden patterns from the CDR, it is recommended to convert tabular data of CDR into graphical one. Different tools are available to convert tabular data into graphical one. One such tool is IBM I2 analyze. This tool will convert the tabular data into graphical one as shown in Figure 3.

Nodes in Figure 3 represent the calling parties. While the frequency of calls are shown through the edges. Node A represents the subject, while nodes B, C, D, E, F and G are the one with whom the patient was in contact. From Figure 3, one can easily interpret that the subject is frequently in contact with party C, with a total of 80 outgoing calls and 39 incoming calls. Patient is in least contact with party G with only 4 outgoing calls and 1 incoming call. While investigating a COVID-19 patient, Figure 3 depicts a vital role to memorize the patient with whom he/she has met. CDRA analysis outcomes will help the patient to reduce the chances to miss anyone with whom the patient has met. In fact all the parties those were in contact with the patient will be contacted. It will be confirmed whether they have met the patient or not. Through CDR of a patient, one can easily improve the accuracy of tracking and tracing the suspect ones. After identifying the people who have met with the patient will be sent into the quarantine. During the process the suspected people will be guided to stay at their homes or remain isolated until the final result of the test is declared. If the test result is positive then the patient will be sent to hospital. Similar procedure will be adapted to all those people to whom the patient has been interacted. In addition to this through CDRA one can easily track the path followed by the patient using cell site detail and latitude and longitude. With the help of cell site data and latitude and longitude of cell site available in Table 2, one can easily plot the map of the COVID-19 patient using Google map. Figure 4 shows the Google map plot with the help of cell site detail given in CDR. This map will be made available to general public and alerts will be generated to avoid this path.

After investigating patient A in the light of CDR graphical links. Let suppose patient A admits that during this period he/she has met with C, D and E. Therefore node C, D and E will be labeled as suspects as shown in Figure 5. The concerned government department will immediately contact C, D and E and will ask them to quarantine them-
TABLE 2

| Date & Time      | A party | B party | Call Type         | IMEI                | Cell Site       | Latitude    | Longitude   |
|------------------|---------|---------|-------------------|---------------------|-----------------|-------------|-------------|
| 08/01/2020 11:13:04 | 1234567890 | 121212121 | Call Outgoing    | 3530030719058      | Plot # 1, Rawalpindi | 33.52292    | 73.23864    |
| 08/01/2020 12:30:02 | 1234567890 | 1313131313 | Call Incoming    | 3530030719058      | Plot # 2, Rawalpindi | 33.5026     | 73.1965     |
| 05/07/2020 14:35:36 | 1234567890 | 1414141414 | Call Outgoing    | 3530030719058      | Plot # 2, Islamabad  | 33.5647     | 73.0367     |
| 05/07/2020 14:35:42 | 1234567890 | 1717171717 | Call Incoming    | 3530030719058      | Plot # 3, Islamabad  | 33.65482    | 73.04976    |
| 05/03/2020 17:16:29 | 1234567890 | 1818181818 | Call Outgoing    | 3530030719058      | Plot # 4, Islamabad  | 33.5695     | 73.13191    |
| 05/04/2020 17:46:14 | 1234567890 | 1919191919 | Call Incoming    | 3530030719058      | Plot # 5, Islamabad  | 33.5795     | 73.1506     |
| 05/05/2020 17:49:33 | 1234567890 | 1212121212 | Call Incoming    | 3530030719058      | Plot # 6, Islamabad  | 33.5796     | 73.1452     |
| 05/06/2020 17:51:23 | 1234567890 | 1414141414 | Call Incoming    | 3530030719058      | Plot # 6, Islamabad  | 33.5796     | 73.1452     |
| 05/06/2020 17:52:35 | 1234567890 | 1717171717 | Call Outgoing    | 3530030719058      | Plot # 8, Islamabad  | 33.569      | 73.1391     |
| 05/07/2020 10:45:24 | 1234567890 | 1919191919 | Call Incoming    | 3530030719058      | Plot # 9, Islamabad  | 33.65482    | 73.04976    |
| 05/07/2020 12:39:40 | 1234567890 | 1616161616 | Call Outgoing    | 3530030719058      | Plot # 10, Islamabad | 33.65482    | 73.04196    |
| 05/07/2020 12:47:11 | 1234567890 | 1515151515 | Call Incoming    | 3530030719058      | Plot # 10, Islamabad | 33.66482    | 73.04196    |
| 05/07/2020 14:35:37 | 1234567890 | 1313131313 | Call Outgoing    | 3530030719058      | Plot # 10, Islamabad | 33.66482    | 73.04196    |

Fig. 5. Identification of coronavirus suspects

Fig. 6. Identification of coronavirus patient

Fig. 7. Graphical interpretation of node D CDR

selves until and unless their corona tests come negative.

Suppose that after testing the suspect ones, only D comes positive then this node will be labeled as patient and the graph will be modified as shown in Figure 6.

Now the entire process will be repeated for node D i.e., CDRA of node D as shown in Table 3. Its graphical interpretation is shown in Figure 7.

Google map of the patient node D will be created with the help of data available in CDR of node D as shown in Figure 8. The map will be made public with an alert to avoid the use of this path for couple of weeks.

After investigating node D, one can easily find out the suspects in the light of CDR graph linkages and path followed by the patient as shown in Figure 7, 8, respectively. Similarly, the same cycle will be repeated to dig out the suspected ones as shown in Figure 9. This process will keep...
TABLE 3
Relevant information contained in patient’s CDR (node D)

| Date & Time       | A party IMEI    | B party IMEI    | Call Type   | Cell Site     | Latitude     | Longitude     |
|-------------------|-----------------|-----------------|-------------|---------------|--------------|--------------|
| 05/01/2020 11:13:04 | 9876543210      | 21121212121    | Call Outgoing | 2110030719058 | Plot # 1, Rawalpindi | 33.66822 | 73.23864 |
| 05/08/2020 12:30:02 | 9876543210      | 31131313131    | Call Incoming | 2110030719058 | Plot # 2, Islamabad  | 33.66822 | 73.23864 |
| 05/08/2020 12:35:36 | 9876543210      | 41141414141    | Call Outgoing | 2110030719058 | Plot # 3, Islamabad  | 33.66822 | 73.23864 |
| 05/09/2020 14:41:29 | 9876543210      | 51151515151    | Call Incoming | 2110030719058 | Plot # 4, Islamabad  | 33.66822 | 73.23864 |
| 05/10/2020 17:16:29 | 9876543210      | 61161616161    | Call Outgoing | 2110030719058 | Plot # 5, Islamabad  | 33.66844 | 73.01676 |
| 05/11/2020 17:32:06 | 9876543210      | 71171717171    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |
| 05/12/2020 17:46:14 | 9876543210      | 81181818181    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |
| 05/12/2020 17:49:33 | 9876543210      | 81181818181    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |
| 05/13/2020 17:51:23 | 9876543210      | 81181818181    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |
| 05/14/2020 17:52:35 | 9876543210      | 81181818181    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |
| 05/15/2020 17:54:24 | 9876543210      | 81181818181    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |
| 05/16/2020 17:47:11 | 9876543210      | 81181818181    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |
| 05/16/2020 14:35:37 | 9876543210      | 81181818181    | Call Outgoing | 2110030719058 | Plot # 6, Islamabad  | 33.66844 | 72.998836 |

3 CONTACT TRACING FRAMEWORK

Since the proposed model is based on CDR, which only works once a suspected person made a call and its log is maintained and tracked. However, in a scenario where a person who is infected, meet to another person without having a phone call or go into the crowd, then the model switches to exposure notifications framework. Exposure notifications system (ENS) previously known as contact tracing is a technique in which two nearby mobile devices exchange random sequence of keys and a log is maintained in a cloud storage to keep record of all the devices that have been remained in physical contact previously. This framework is developed by the joint efforts of Google and Apple. Most importantly this technique protect user’s privacy and security.

3.1 How it Works?

People all around the world continuously working on developing apps based on exposure notification system to help them in tracing the people that remained in contact with infected patients of Covid-19. Once an individual selects this model, a unique random key of 4 digits generates on that mobile device. This key is continuously changed after 10 to 20 minutes so that to ensure it cannot be used to recognize the location. A phone with ENS based app comes into contact with another device embedded with same app, both phones exchange these random keys with each other through bluetooth low energy (BLE) as shown in Figure 10. This not only ensures the privacy but also improves the overall power consumption.

User’s phone periodically checks all the random keys that have been maintained to identify a patient tested positive with COVID-19. If there is a match exists, user and concerned government department receive exposure notification on its phone. In this way one can overcome the limitation of CDR when a person comes into the contact with another without having a phone call.

3.2 User Privacy

Privacy always remains at first. In exposure notifications a user may allow it anytime and can disallow with its own consent. As discussed above, exposure notifications do not track location of your device and only works based on BLE. Exposure notification matching only occurs on the device and this framework does share identity with other people, Google, or Apple. Concerned authorities may ask for contact details but only with the consent of user.
3.3 Case Study 2

In order to integrate the information and interpret the outcome of the proposed model, one complete case study based on hypothesis is discussed below. Two people named A and B go to a marketplace at the same time and place. They both have contact tracing app installed. After the shopping, they wait for 10 minutes in check out queue. During this time interval their phones exchange random keys to each other and store information. Thus they know each other that they have remained in contact for a while as shown in Figure 11.

After few days, person A feels symptoms of COVID-19 and consults a doctor. Unfortunately he is diagnosed and tested positive for COVID-19. Person A opens up his app and verifies his documentation provided by health care provider. He simply taps the button and declares himself COVID-19 patient as shown in Figure 12. This declaration automatically uploads on storage cloud once Person A declares himself COVID-19 patient. After that Person B phone downloads beacons from all those people who have been diagnosed with COVID-19. In this way Person B can quarantine himself or can perform test for COVID-19. Later if person B diagnoses with COVID-19 he will perform same steps as person A to declare himself COVID-19 patient as shown in Figure 13. Similarly, if C was in contact with number of people without a phone call then the entire log will be maintained in the same manners as shown in Figure 14.

Hence this model is capable to overcome the tracking limitation of CDR when call is not made. This paper presents an efficient and robust framework to track COVID-19 patients based on CDR and ENS. This model will help the governments to track, trace and terminate coronavirus cases by exploring the entire network of COVID-19 patients.
The capabilities of proposed model is summarized in Table 4

4 Conclusion

Since there is no vaccine available to cure this pandemic disease, therefore, the best solution is to minimize its spread. This paper has proposed a novel model to control the widespread of coronavirus using CDRA and contact tracing. The model proposed in this paper can help to track and trace the COVID-19 patients. Through this analysis, one can easily extract the information with whom the patient has remained in contact. So with CDRA analysis one can easily track the COVID-19 patients and suspects. Through CDRA, we can easily track the path followed by patients for the last couple of weeks and alerts can be generated for general public to avoid that path. The case study section has thoroughly discussed how to implement this model. One of the main limitations of CDRA is when a suspect goes into the crowd without making a phone call. In order to overcome this limitation contact tracing based on ENS is used in parallel to track the path traversed made by the patient without call history. The information extracted with the help of this hybrid model can easily minimize the wide spread of coronavirus disease. The potential outcomes of the proposed model are tracking of COVID-19 patients, monitoring of isolated patients, tracking of suspected ones and inform the mass about the safest path to use. It is therefore recommended that the proposed model can be adopted by the government to control the widespread of coronavirus disease. In future, the authors are planning to include different social networks data such as Facebook, WhatsApp, Google, Telegram and Instagram in the proposed model. This will further improve the tracking mechanism of COVID-19 patients.

References

[1] Jane, Y.; Lucy, U.; Zadravecz; Frank; Cariello; Paloma; et al. Novel coronavirus 2019 (COVID-19): emergence and implications for emergency care. *Journal of the American College of Emergency Physicians Open* 2020.

[2] Park, M.; Cook, A.R.; Lim, J.T.; Sun, Y.; Dickens, B.L. A systematic review of COVID-19 epidemiology based on current evidence. *Journal of Clinical Medicine* 2020, 9, 967.

[3] Norman, G. The Once and Future Myths of Medical Education. John Wiley & Sons, 2007.

[4] Hassan, S.A.; Sheikh, F.N.; Jamal, S.; Ezeh, J.K.; Akhtar, A. Coronavirus (COVID-19): a review of clinical features, diagnosis, and treatment. *Cureus* 2020, 12.

[5] Nicola, M.; Alsafi, Z.; Sohrabi, C.; Kerwan, A.; Al-Jabir, A.; Isosifidis, C.; Agha, M.; Agha, R. The socio-economic implications of the coronavirus and COVID-19 pandemic: A review. *International Journal of Surgery* 2020.

[6] Sahoo, P.P.; Rath, S. Potential Impact of Corona Virus on Agriculture Sector. *Biotica Research Today* 2020, 2, 64–65.

[7] Sohrabi, C.; Alsafi, Z.; O’Neill, N.; Khan, M.; Kerwan, A.; Al-Jabir, A.; Isosifidis, C.; Agha, R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery* 2020.

[8] Rana, S. An unprecedented footprint of covid19 on education sector: A review on impact and measures. *Purakala with ISSN 0971-2143 is an UGC CARE Journal* 2020, 31, 548–551.

[9] Schiffbauer, M.T. Western Balkans Regular Economic Report, No. 17, Spring 2020: The Economic and Social Impact of COVID-19, 2020.

[10] Cavallo, J.J.; Forman, H.P. The economic impact of the COVID-19 pandemic on radiology practices. *Radiology* 2020, p. 201495.

[11] Hoque, A.; Shikha, F.A.; Hasanat, M.W.; Arif, I.; Hamid, A.B.A. The effect of Coronavirus (COVID-19) in the tourism industry in China. *Asian Journal of Multidisciplinary Studies* 2020, 3, 52–58.

[12] Fetzer, T.; Hensel, L.; Hermle, J.; Roth, C. Coronavirus perceptions and economic anxiety. *arXiv preprint arXiv:2003.03848* 2020.

[13] Parnell, D.; Widdop, P.; Bond, A.; Wilson, R. Crime investigation using Call Data Records (CDR) through Big Data technology. 2017 International Conference on Nascent Technologies in Engineering (ICNTE). IEEE, 2017, pp. 1–5.

[14] Ansari, B.; Ganjoo, M. Impact of Covid-19 on Advertising: A Perception Study on the Effects on Print and Broadcast Media and Consumer Behavior. *Purakala with ISSN 0971-2143 is an UGC CARE Journal* 2020, 31, 52–62.

[15] Bradbury-Jones, C.; Isham, L. The pandemic paradox: the consequences of COVID-19 on domestic violence. *Journal of clinical nursing* 2020.

[16] Armitage, R.; Nellums, L.B. COVID-19 and the consequences of isolating the elderly. *The Lancet Public Health* 2020, 5, e256.

[17] Horak, R. *Telecommunications and data communications handbook*. John Wiley & Sons, 2007.

[18] Khan, S.; Ansari, F.; Dhalvelkar, H.A.; Computer, S. Criminal investigation using Call Data Records (CDR) through Big Data technology. 2017 International Conference on Nascent Technologies in Engineering (ICNTE). IEEE, 2017, pp. 1–5.

[19] Kumar, M.; Hanumanthappa, M.; Kumar, T.S. Crime investigation and criminal network analysis using archive call detail records. 2016 Eighth International Conference on Advanced Computing (ICoAC). IEEE, 2017, pp. 46–50.

[20] Google Exposure Notification API launches to support public health agencies. 2020.

[21] Apple and Google Contact Tracing – Bluetooth Specification. 2020.