

RESEARCH ARTICLE

Population data mobility retrieval at territory of Czechia in pandemic COVID-19 period

Jan Platos¹  | Pavel Kromer¹  | Miroslav Voznak² | Vaclav Snasel¹

¹Department of Computer Science, FEECS, VSB-Technical University of Ostrava, Ostrava, Czech Republic

²Department of Telecommunications, FEECS, VSB-Technical University of Ostrava, Czech Republic

Correspondence

Jan Platos, Department of Computer Science, FEECS, VSB-Technical University of Ostrava, 17. listopadu 2172/15, 70800 Ostrava, Czech Republic.

Email: jan.platos@vsb.cz

Funding information

Vysoká Škola Báňská - Technická Univerzita Ostrava, Grant/Award Number: SGS 2020/108

Abstract

This article describes the methodology and the possibilities of collecting operation data in a mobile network provider. First, the architecture and the principles used in the system are described. The precision analysis of the population commuting in the region and during the pandemic and nonpandemic times. Moreover, several ideas about further utilization of the data will be formulated and described. Finally, a graph-based approach that describes the creation of the community structure between the people and the means of its analysis.

KEYWORDS

COVID-19, data analysis, mobile networks, population mobility

1 | INTRODUCTION

The tremendous penetration of mobile phones in the population brings many new challenges and many difficulties. The traffic that is generated by the usage of mobile phones is enormous due to colossal internet traffic. The service traffic is also generated by logging into the system, beginning and end of the conversation, and other events. Such information may be utilized to gather essential characteristics of the masses' behavior, such as a population in a city and/or country. The information for the person tracking violates law regulation, such as General Data Protection Regulation and similar in each country. The utilization of the anonymous aggregated data brings a new insight into the behavior patterns and may analyze the specific situation.

Many real-life situations may benefit from such anonymous aggregated data. The first case is the replacement of the personal survey of the occupancy in ground transport. In 2014, we defined a methodology¹ that was certified by the Ministry of Transportation of the Czech Republic, where the means of transport and behavior of the citizens were categorized and classified. Moreover, the methodology defines the categories of citizens according to their behavior during the day.

The most important categories are commuter and noncommuter. A commuter is a person who begins in one location, then travels to another location where he spends some time, and then returns. Such a person is usually a worker who commutes to work from its home. The commuting path may be tracked and analyzed to identify the primary transport nodes, bottlenecks, etc. The noncommuter stays the whole day in a location where he started the day and never left it.

Such a complex system that can track anonymous masses commuting behavior and paths that were used may be utilized to track the contact between individuals while maintaining the system's anonymous property.

2 | MOBILE PHONE LOCATION DATA ANALYSIS

Mobile (cell) phone data has been used to provide useful information about human mobility and traffic patterns for almost two decades.²⁻¹³ Mobile (cellular) network operators record the location of a mobile phone when it actively uses the network (i.e., makes or receives a call, uses mobile

Internet). The information is stored primarily for billing purpose² but it can be used as a latent source of aggregate spatiotemporal information about static and dynamic human mobility and traffic patterns.^{4,7,10,13}

Mobile phone data can be in this way exploited to study travel and commuting times,^{2,5,9} to provide information about congestions and traffic incidents,^{2,9} and to detect origin-destination flows.^{2,5,11} Low-level location data requires the transformation to spatiotemporal trajectory (path) information. Such information can then be used by high-level applications such as early warning, environmental monitoring (e.g., air pollution sensing), traffic planning,^{3,4} region-level travel demand estimation,⁵ sociological (sociogeographical) research.⁶

A special attention has been paid to the study of the relationship between mobile phone data and human mobility and public transport.^{8,11,12} New methods to infer from mobile phone location data traffic measures (e.g., vehicle flows and densities on freeways and motorways)⁸ and top-level information such as home location^{6,10,13} and activity radius (space),^{6,10} travel demand,^{5,11} anchor points (stay locations) of daily travels, and activity types¹³ have been recently introduced.

The methods include various data preprocessing techniques such as low-pass filtering,⁵ noise removal and thresholding¹³ and so on. Then, data processing and mining in the form of cluster analysis^{3-5,8,10,13} and outlier detection,^{7,13} topology construction,^{3,4} signal analysis (dynamic timewarping),⁷ density analysis,^{6,11} motif extraction,¹³ and, for example, visual analytics¹² are employed.

In addition, mobile phone location data has been successfully used in context of epidemiology.¹⁴⁻¹⁶ It can contribute to the modeling and estimation of human contact dynamics and disease spread.¹⁴ Mobile phone data and the extracted communication and mobility patterns have been used to study the spatial structure and variation of the HIV epidemics and to identify disease hot spots.¹⁵ Together with social network data, it has shown a great potential for disease tracking and outbreak predictions, as demonstrated on the 2010 cholera epidemic in Haiti.¹⁶

The coronavirus 2019–2020 pandemic is undoubtedly the biggest challenge digital¹⁶ and mathematical¹⁴ epidemiology faces to date. Until the introduction of an efficient vaccine and/or COVID-19 treatment (unavailable as of June 2020), social distancing and economy shutdown are the only measures available to slow-down the epidemics. Extensive testing and tracing and highly focused epidemiological procedures can lower the enormous costs associated with the global coronavirus suppression measures. Mobile phone data can be used to tackle COVID-19 pandemic in several ways:¹⁷ by helping to establish accurate situational awareness and by evaluating the impact of various interventions performed in order to contain the disease. This can be achieved by, for example, the estimation of origin-destination flows, location and stay (hot spot) modeling, approximation of contact matrices, and so on.

3 | SPATIAL-TEMPORAL MOBILITY OF THE POPULATION

The utilization of mobile networks for the population mobility analysis opens an entirely new area of the investigation. First of all, the system needs to be described.

3.1 | System architecture

The system's whole architecture is based on the architecture and the processes behind the wireless communication networks. The standard architecture is shown in Figure 1.

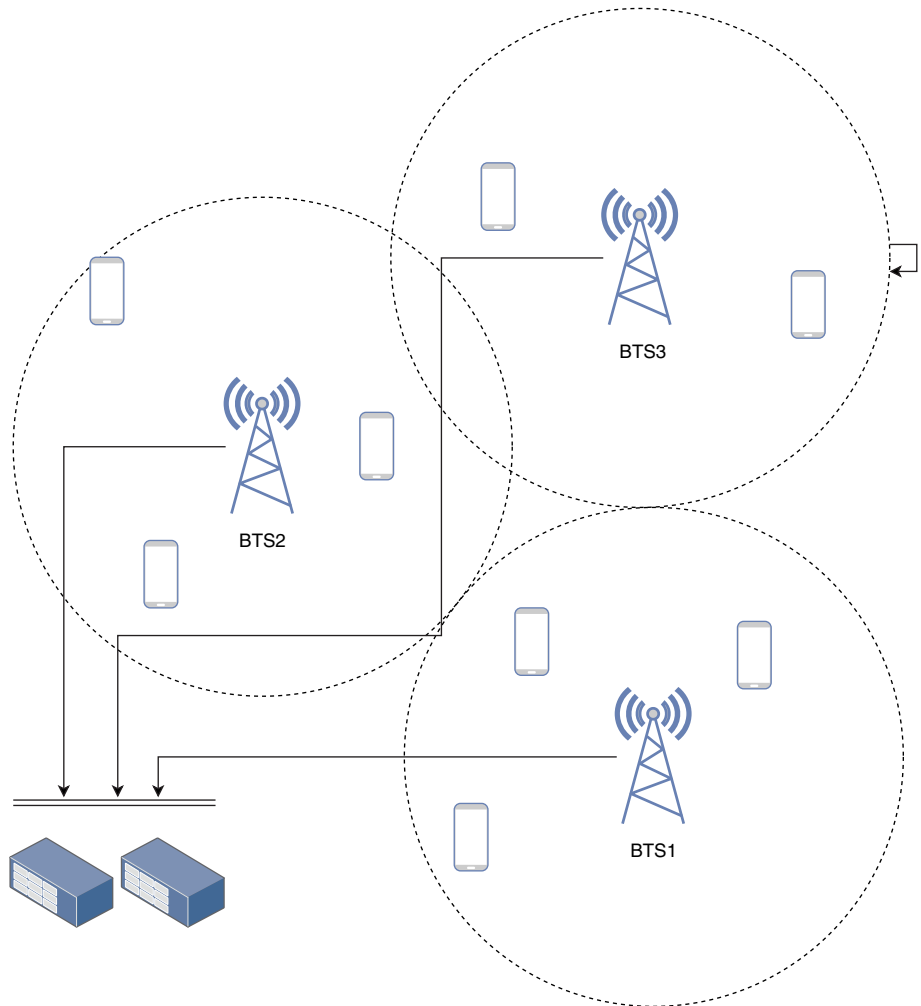
Each cell phone communicates with the nearest Base transceiver station (BTS), which is the communication network's main building block. Each BTS is connected with the servers of the communication provider. The location of the BTS in the real world is based on many attributes. The main attributes are:

- predicted the number of devices connected at the same time,
- complexity of the landscape,
- number of buildings and reflection caused by buildings around the station,
- directions toward the main location of the devices (such as subway exits, etc.).

The base stations' placement then creates the topology of the telecommunication network, which may be utilized for many other applications other than pure communication. The first application that is used by the phone operating system providers is the improvement of the phone location based on the triangulation between base station that is used as a correction to the GPS signal.⁹

Each mobile device that is connected to the telecommunication network generates a vast footprint during device usage. Each operation that the device performs, such as short message service, audio, video calls, and data transactions, generates a communication with the network that is tracked and stored in logs. Many countries have legislative regulations that define how much information needs to be stored for the police and other purposes. The Czech Republic defined that this information has to be stored for 3 months.

FIGURE 1 Architecture of the cell phone architecture



3.2 | Spatial-temporal mobility analysis

Analysis of the population's mobility may be gathered using many different systems, for example, GPS locations, debit cards, Internet of Things, and many others. Mobile networks have the main advantage in penetrations. The Czech Republic registers more SIM cards than is the population of the whole country. Such penetration leads to ideal conditions for utilizing this data in the analysis of the population's mobility pattern.

The data analysis is done in a single day cycle from 12:00 a.m. to 11:59 p.m. for each day. The first 5 h are not very important, and we may expect that this time is a quiet night period. At 5:00 a.m. in the morning, start a day period, where each person/device's behavior is investigated. The location where the device is at five is taken as a starting home location. The location is tracked for the whole day, and it is expected that the device will move from the starting location to other places. In the majority, the movement leads to the working location. The sequence of the cells and BTS used during the movement is carefully tracked and evaluated for precise mapping of the location. Of course, the movement usually starts much later than at five. However, the 5:00 a.m. is the threshold, investigated, and defined in the Methodology to be a good time for division the night and day periods. The tracking continues during the whole day until the device ends in a location in the evening and night hours. When the location is the same as it was in the morning, we consider this location a full home location. Otherwise, the person changes its night location for obvious reasons such as business trips, etc. Such behavior has limitations and exceptions, which are mostly visible on Mondays, Fridays, and Saturdays. On Monday and Friday, many people travel from the weekend location to the work location and back. Fridays and Saturdays are also days for event trips with friends etc.

When all devices are investigated, we receive the heat map of locations and stations where people appeared and its time distribution during the day. Such information may be aggregated to receive the overall statistics and its development during the day. The system parameters strictly limit the granularity of the statistics. The optimal interval is hours because the hour is used to refresh the communication between the device and the network when no other activity is performed. Due to fast development in the device's quality, shorter intervals may be selected, such as 30 and 15 min. The 4G networks use a shorter interval for synchronization and *εi-am-alive* messages between the device and the network. The 5G networks will bring a new problem because they use adaptive syncing, where the interval is not defined as fixed.

4 | UTILIZATION OF THE GATHERED DATA

The data aggregated during the day on the individual base stations must be grouped for the higher territorial unit. On a single pylon, the standard behavior contains several base station for each technology and several different technologies. The aggregation process may be defined as a sequence of the steps that follow each other to bring a more globalized view of the data. The details are shown in Figure 2. When the data are aggregated, there is much application that maybe build on it.

4.1 | Behavior patterns in population mobility

When all the aggregated data are collected for a defined time interval, a behavior analysis may be performed. There are two basic views of the analysis. The first is based on the analysis of the occupancy of a place. The second is the population flow in time.

4.1.1 | Static analysis of the people location

The occupancy or the utilization of the concrete base stations is an analysis relevant for evaluating the concentration of people in defined places. Such information may be utilized for:

- planning of public transportation services,
- crowd avoidance,
- security and safety analysis,
- urban planning,
- phenomena detection,
- unique person visits.

For each place - base station location, city part, city, region, we can count the number of unique devices/persons each time unit. This information is not very usable until the differences between dates are analyzed.

As may be seen from Figure 3, the difference between the long-term average and the current situation. The high decrease of the commuters and even higher increase in noncommuters indicates that the strange situation has happened because very few people come into the area and even fewer of them leave the area during the day. Of course, the strange situation is a COVID-19 pandemic and the almost lock-down of the whole country—more about this topic in the next section.

The long-term data collection in a specific location may be utilized to predict the area's occupancy in the future. Such information may be utilized to plan public transport and gather the number of persons for future calculation. An example of such prediction is depicted in Figure 4. As may be seen, the prediction may be made with defined confidence intervals.

4.1.2 | Dynamic analysis of the people location

The population flow in time is valuable information for:

- public transport planning in the inter-city level,
- occupancy of the transport vehicle measurement,

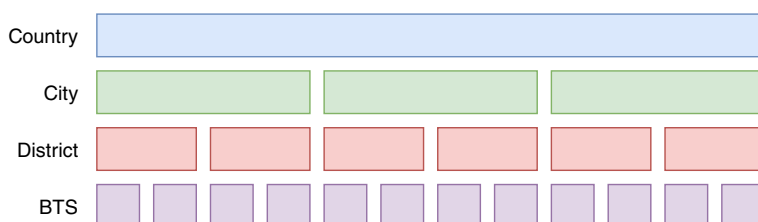


FIGURE 2 Aggregation process for the day cycle

FIGURE 3 The differences in percents of the commuters and non-commuters between March 15 and May 30 in Prague 1 city district

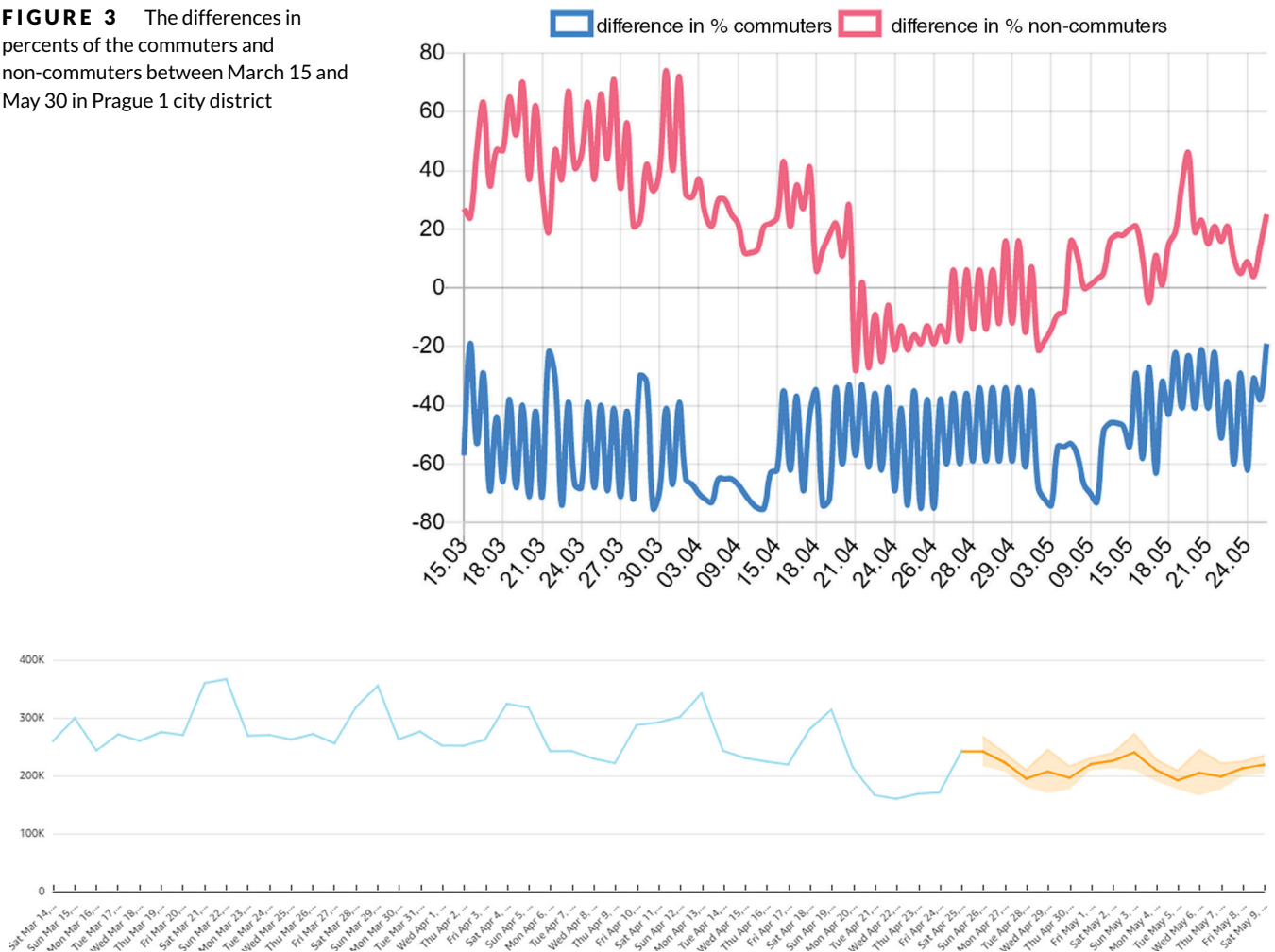


FIGURE 4 Prediction for 14 days of the cumulative amount of persons

- road capacity utilization,
- traffic control,
- accident detection.

The occupancy of the public transport services' vehicles was the original idea extended into many other applications. For measurement of this occupancy, precise information of the road network, station locations, vehicle movement logs is necessary.

Figure 5 depicts the relations and number of people that travel between two cities along the trajectory of the public transport service concerning the source and the destination of the people during a day. The number in the edges depicts the comparison between the known number investigated by the Czech Statistical Agency using the survey analysis and the number computed from the data. As may be seen, the dynamic analysis can deliver information about the transport patterns. Because we can track the location and the speed of transportation and approximate the position, we can detect the type of vehicles used in transportation. With the time-tracked GPS position of the public transport vehicles, we may detect people who follow the same pattern as the vehicles, which means they travel by the vehicle. If a person follows the same trajectory, it uses the same roads, but the cars travel at a different speed with no stops on these short trips. The system is then able to distinguish between public transport users and drivers. Even though many factors limit precision, it can achieve very high precision. Moreover, such information may be utilized to collect information about traffic jams, accidents, etc. The measurement of speed can also detect faster vehicles such as emergency helicopters that travel in the same area.

The direct utilization of the data may also bring a new application to personal transport. Especially during the infection disease period, a person wants to avoid large crowds during their transport. The person flow data may help predict crowded places and calculate the path between the source and destination places with the smallest probability of crowded surroundings. Such applications appeared in the COVID Pandemic in many countries and cities.

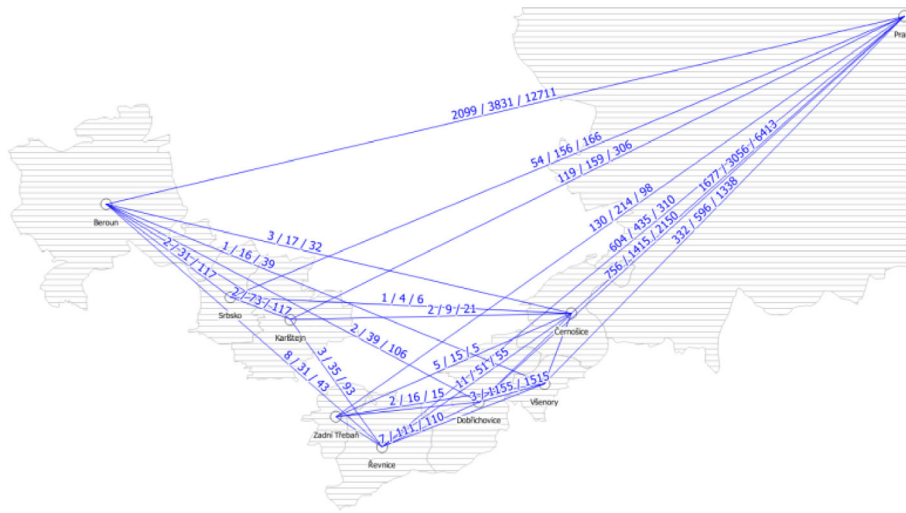


FIGURE 5 Relations of commuters between different cities/stations from city of Beroun to Prague¹

4.2 | Graph-based applications

The detection of a device in a base station generates a vast stream of data from the whole network. Due to the time dependency, the data may be represented as time-dependent graphs. Such a graph, which is demonstrated in Figure 6, may utilize the information from the temporary closeness of the devices/persons to create a probability network of the local neighborhood connections. The figure shows on the left that people change the location during the time and the occupancy on three base stations. The groups of people who appear in each time step on the BTS-B are highlighted. The corresponding graph is depicted on the right. As may be seen, each time step generates cliques that evolve during the time. According to these time slices, a probability graph may be constructed and used to analyze the relationships between persons. Such a graph reflects the probability the person visited a place covered by the BTS.

Such graphs may also be constructed from the logs of the GPS positions, but it requires the persons' cooperation and enabling of such functionality. When a graph is constructed, many different measures may be used for distance measurement to gather the similarity between users based on their mobility during the day. Among others, the shortest path¹⁸ describes the first level distance between nodes. As a more complex distance, a Close Trail Distance¹⁹ measure may be utilized. The behavior of the persons may then be analyzed using the clustering coefficients.²⁰ The behavior patterns may be compared based on the historical records and compared the situations with reference and precisely analyzed situations.

4.3 | Possible applications

The essential aspect of the data is its security. The first thing that needs to be set is that the analysis and statistics are created from the providers' anatomized data and fully anonymized daily. Therefore, the system cannot identify anybody directly from the data. Moreover, the person cannot be identified from its behavior.

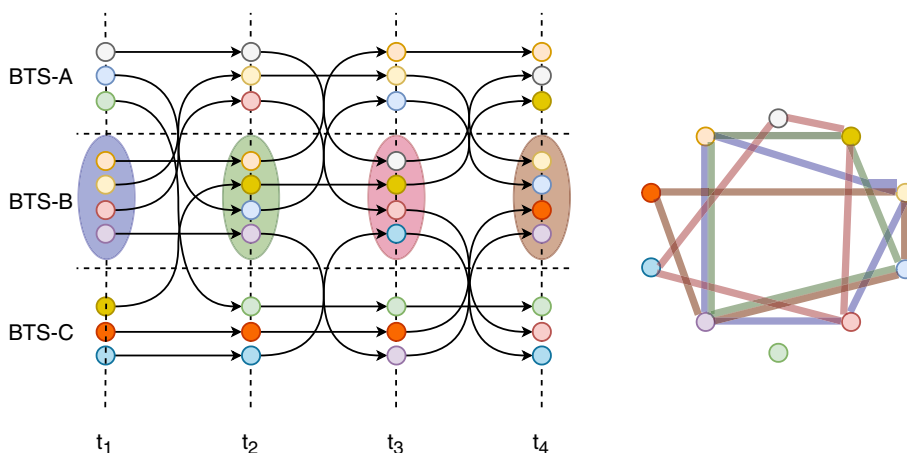


FIGURE 6 Example of the network model based on the collected data

The security issues mentioned in the previous paragraph prevent the use of this data to direct mapping the location of the current people. This is a good thing because mentioning this methodology in the major TV news show leads to a concern of the general public and the personal freedom activist. On the other hand, the methodology may be used in other COVID-related applications. The Czech Government uses these data to monitor the citizens' behavior related to their general mobility and the increase of the number of noncommuters, that is, people who do not commute to the job. The data are used on a day-by-day basis, and the government uses this to check the regulations' compliance.

The compliance with the people quarantine regulation applied to the COVID-positive persons may not be used the anonymous data. However, the government may grant the police, under the state of emergency, to use the nonanonymous data to monitor the quarantined people's real movement. Such data were used by the Israel secret service to monitor Israel citizens during the first wave of the COVID pandemic. However, the practice was forbidden by the constitutional court. Many government applications use very similar approaches to track citizen mobility and contacts. Unfortunately, such an application did not attract many users due to controversial or unclear personal data handling. The new applications based on the technology developed by the Google and Apple company, as the primary cell phone operating system providers attract a higher number of users, and the newly designed application, bring a completely new contact tracking level.

5 | CONCLUSION

In this article, the system for processing the massive data from the telecommunication networks collected by the providers and stored under the national regulations were described. Several applications concerning the nature of the data were presented. Data collection brings completely new problems concerning the General Data Protection Regulation in the EU, national regulations, and other laws and restrictions. An aggregated data without any identification may be analyzed and the usable output produced with a well-defined methodology. The Czech Ministry of Transportation certified the methodology. Other applications include population mobility during the time interval, comparison of the mobility patterns and changes, and anomaly detection. Moreover, several measures for data analysis were summarized.

The mentioned applications need a vast computational power available in the data centers of the network providers. Such analysis will create an entirely new market for the providers. Using methodologies like the mentioned in these articles, the network providers will create an entirely new market with the specialized analysis that is impossible using the classical scheme and data sources.

ACKNOWLEDGEMENTS

This work was supported by SGS, VŠB – Technical University of Ostrava, Czech Republic, under the grant No. SP2020/108 “Parallel processing of Big Data VII”.

ORCID

Jan Platos  <https://orcid.org/0000-0002-8481-0136>

Pavel Kromer  <https://orcid.org/0000-0001-8428-3332>

REFERENCES

1. Voznak M, Hylmar J, Blagodarny D, et al. *Specific Method of Passenger Handling and Number of Transported Passengers (in Czech)*. Technical Report. Poruba, Ostrava: VSB-Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science; 2016.
2. White J, Quick J, Philippou P. The use of mobile phone location data for traffic information. Paper presented at: Proceedings of the 12th IEE International Conference on Road Transport Information and Control, 2004. RTIC 2004; 2004:321-325.
3. Bayir MA, Demirbas M, Eagle N. Discovering spatiotemporal mobility profiles of cellphone users. Paper presented at: Proceedings of the 2009 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks Workshops; 2009:1-9.
4. Bayir MA, Demirbas M, Eagle N. Mobility profiler: a framework for discovering mobility profiles of cell phone users. *Pervasive Mob Comput*. 2010;6(4):435-454. Human Behavior in Ubiquitous Environments: Modeling of Human Mobility Patterns. <https://doi.org/10.1016/j.pmcj.2010.01.003>.
5. Calabrese F, Ratti C, Lorenzo GD, Liu L. Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Comput*. 2011;10(04):36-44. <https://doi.org/10.1109/MPRV.2011.41>.
6. Phithakkitnukoon S, Smoreda Z, Olivier P. Socio-geography of human mobility: a study using longitudinal mobile phone data. *PLOS One*. 2012;7(6):1-9. <https://doi.org/10.1371/journal.pone.0039253>.
7. Yuan Y, Raubal M. Extracting dynamic urban mobility patterns from mobile phone data. In: Xiao N, Kwan MP, Goodchild MF, Shekhar S, eds. *Geographic Information Science*. Berlin, Heidelberg/Germany: Springer; 2012:354-367.
8. Gao H, Liu F. Estimating freeway traffic measures from mobile phone location data. *Europ J Operat Res*. 2013;229(1):252-260. <https://doi.org/10.1016/j.ejor.2013.02.044>.
9. Becker R, Cáceres R, Hanson K, et al. Human mobility characterization from cellular network data. *Commun ACM*. 2013;56(1):74-82. <https://doi.org/10.1145/2398356.2398375>.
10. Xu Y, Shaw SL, Zhao Z, Yin L, Fang Z, Li Q. Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*. 2015;42(4):625-646. <https://doi.org/10.1007/s11116-015-9597-y>.
11. Demissie MG, Phithakkitnukoon S, Sukhivil T, Antunes F, Gomes R, Bento C. Inferring passenger travel demand to improve urban mobility in developing countries using cell phone data: a case study of senegal. *IEEE Trans Intell Transp Syst*. 2016;17(9):2466-2478.

12. Di Lorenzo G, Sbodio M, Calabrese F, Berlingiero M, Pinelli F, Nair R. AllAboard: visual exploration of cellphone mobility data to optimise public transport. *IEEE Trans Vis Comput Graph*. 2016;22(2):1036-1050.
13. Jiang S, Ferreira J, Gonzalez MC. Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore. *IEEE Trans Big Data*. 2017;3(2):208-219.
14. Hashemian MS, Stanley KG, Knowles DL, Calver J, Osgood ND. Human network data collection in the wild: the epidemiological utility of micro-contact and location data. Paper presented at: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. IHI '12. Association for Computing Machinery; 2012:255-264; New York, NY. <https://doi.org/10.1145/2110363.2110394>
15. Brdar S, Gavrić K, Culibrk D, Crnojević V. Unveiling spatial epidemiology of HIV with mobile phone data. *Sci Rep*. 2016;6(1):19342. <https://doi.org/10.1038/srep19342>.
16. Bates M. Tracking disease: digital epidemiology offers new promise in predicting outbreaks. *IEEE Pulse*. 2017;8(1):18-22.
17. Oliver N, Lepri B, Sterly H, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*. 2020;6 (23):1-7. <http://dx.doi.org/10.1126/sciadv.abc0764>.
18. Knuth DE. A generalization of Dijkstra's algorithm. *Inf Process Lett*. 1977;6(1):1-5. [https://doi.org/10.1016/0020-0190\(77\)90002-3](https://doi.org/10.1016/0020-0190(77)90002-3).
19. Snasel V, Drazdilova P, Platos J. Closed trail distance in a biconnected graph. *PLOS One*. 2018;13(8):1-12. <https://doi.org/10.1371/journal.pone.0202181>.
20. Prokop P, Snasel V, Drazdilova P, Platos J. Clustering and Closure Coefficient Based on k-CT Components. *IEEE Access*. 2020;8 101145-101152. <http://dx.doi.org/10.1109/access.2020.2998744>.

How to cite this article: Platos J, Kromer P, Voznak M, Snasel V. Population data mobility retrieval at territory of Czechia in pandemic COVID-19 period. *Concurrency Computat Pract Exper*. 2020;e6105. <https://doi.org/10.1002/cpe.6105>