On Use of Big Data for Enhancing Network Coverage Analysis

Ömer Faruk Çelebi, Engin Zeydan, Ömer Faruk Kurt, Ömer Dedeoğlu, Ömer İleri, Burak Aykut Sungur, Ahmet Akan, Salih Ergüt

***Abstract*— Proliferation of data services has made it mandatory for operators to be able identify geographical regions with 3G connectivity discontinuity in a scalable and cost-efficient manner. The currently used methods for such analysis are either costly - such as in drive tests, partly unreliable - such as in network simulation approaches, or are not precise enough- such as in base station key performance indicators (KPI) based approaches. In this paper, towards addressing these inadequacies, we propose a 3G coverage analysis method that makes use of "big data" processing schemes and the vast amounts of network data logged in mobile operators. In the proposed scheme, the BSSAP mobility and radio resource management messages between the BSS and MSC nodes of the operator network are processed to identify inter-technology handovers from 3G (WCDMA) access to 2G (EDGE, GPRS, GSM). Demonstrative examples show that the proposed mechanism produces accurate and precise results, outperforming the base station KPI-based approach.**

***Index Terms*— big data, cellular networks, coverage, Hadoop**

I. INTRODUCTION

Digitization of content and emergence of wireless communication technologies which enable data transmission has placed mobile data services as an important revenue source for mobile operators. This trend has further been emphasized after the emergence of third generation (3G) networks (Wideband-CDMA in Europe), which have provided substantial gains in data speeds compared to 2G (EDGE, GPRS) technologies. On the customer side, the technological availability of high data rate connections has transformed the growing "demand" for mobile data to a "need" for "any-time any-where" connectivity. Therefore, providing ubiquitous and high quality (QoS) mobile data services, through data connectivity, has become priority for mobile operators in most parts of the world.

A major challenge that needs to be addressed by the operators in this respect, is the discontinuity in 3G connectivity experienced by mobile users. Such discontinuity is experienced when the user is forced to handover from 3G connectivity down to 2G, possibly due to loss of signal quality. Such vertical handovers result in less efficient use of the spectrum - through low data rate connections - thus reducing the revenues from data services and also cause imperception regarding service quality on the customer side. It is therefore of utmost importance for operators to identify geographical regions with discontinuities in 3G connectivity.

Identifying regions with 3G service discontinuity might prove to be challenging, as the coverage area for a typical 3G transmitter does not only depend on signal path loss phenomena, but also on the overall planning and dimensioning of the surrounding 2G and 3G base stations. Furthermore, the coverage area is not always steady, as it is also influenced by the time varying band and channel characteristics as well as the instantaneous capacity demand in the cell, particularly in the case of 3G access. The traditional methods of identifying regions with 3G service discontinuity, such as drive-around tests and network simulations do not address the dynamics outlined above, and often are costly. Other methods making use of the base station KPI parameters that track 3G to 2G handovers often do not provide precise enough results, as they only include the signaling information that belongs to "connected mode" of mobile devices, ignoring the "idle mode" dynamics that correspond to time intervals in which the mobile is silent. Therefore, there is need to develop coverage analysis methodologies that can accurately and precisely pinpoint service discontinuity in a cost efficient manner. Using big data processing platforms in telecommunication field is studied in different papers in the literature [1] [2] [3] [4] [5] [6] [7].

In this paper, inspired by the use of "big-data" tools and concepts in diverse fields of study - such as customer-centric marketing, we propose a methodology making use of the vast amounts of signaling data that is constantly produced by the mobiles and fed back to the operator. In the context of a mobile operator, "big data" refers to the collection of vast quantities of data that are being generated by mobile devices such as signaling and location data, web site visits, and several million CDRs (Call Detail Records) produced on a daily basis. In the scope of this paper, we propose to process the network logs for mobile uses, corresponding to both their idle and connected mode signaling data using the Hadoop platform [8], to come up with dynamic 3G service discontinuity maps. More specifically, mobility and radio resource management signaling data (BSSAP messages) collected by the probes attached to the relevant network switches are used. Through such data, the inter radio access technology handovers (IRAT HOs) from 3G to 2G for mobile users are mapped to geographical locations of the users to identify the 3G service discontinuity areas. Experimental results show that the service discontinuity areas achieved through this approach match the drive test results. Moreover, the approach provides more accurate results and provides a more detailed coverage map than the aforemen-

Ö. F. Çelebi, E. Zeydan, Ö. F. Kurt, Ö. İleri, B. A. Sungur, and S. Ergüt are with AveaLabs, and Ö. Dedeoğlu, A. Akan are with Radio Network Planning Department at AVEA İletişim Hizmetleri A.Ş. İstanbul, Turkey. (e-mail: omerfaruk.celebi@avea.com.tr, engin.zeydan@avea.com.tr, omerfaruk.kurt@avea.com.tr, omer.ileri@avea.com.tr, burakaykut.sungur@avea.com.tr, salih.ergut@avea.com.tr, omer.dedeoglu@avea.com.tr, ahmet.akan@avea.com.tr)

tioned methods based on tracking base station KPIs. The increase in accuracy is caused by the fact that such logs include both the "idle" and "connected" mode signaling data, as opposed to the latter alone. The increase in precision can be attributed to the fact that BSSAP messages provide destination information regarding a hand over, unlike the KPI parameters in a base station. The proposed system thus constitutes a scalable, accurate, and precise approach that captures the time-dynamicity of coverage analysis in a cost efficient manner.

Besides the specifics and benefits of the analysis approach that it advocates, this paper also constitutes a contribution in the sense that it provides a nice example of combining results from the two seeming disjoint areas of research; namely the "big data" processing schemes and coverage analysis tools. The mentioned fields have often been elaborated on separately, especially in the wireless cellular network domain. However, some of the results of one study area can readily be applied to the other, making it more powerful and efficient, as illustrated here.

This paper is organized as follows: Section II provides the background for the proposed approach by providing a brief overview of traditional coverage analysis tools and of the concept of "big data". The section also details the proposed big data based coverage analysis method. Section III discusses the simulation results and finally Section IV gives the conclusions and future work.

II. Big Data Based Coverage Analysis

Establishing continuous 3G connectivity is important for operators, in achieving their aim of providing ubiquitous and high quality services for their customers. The undesired vertical handovers from 3G access are often results of falling RSSI (received signal strength indicator) or SNR (signal-to-noise-ratio) values below a certain threshold. In the case of 3G access, the handovers are also triggered by instantaneous demand for excess capacity.

Most coverage analysis methods considered in literature are based on drive tests and costly field measurements [9] [10]. In 3GPP standardization effort, this work item goes by the name of "minimization of drive tests" (MDT) [11]. In dynamic environments (due to population change, new base station deployments, the use of new services, outages, etc), new driving tests and field measurements need to be conducted on a periodic basis to have up-to-date info regarding the coverage areas and potential handovers likely to be triggered across technologies. Several other methods rely on statistical tools [12]. In these methods, Monte-Carlo simulations are utilized to mimic real network conditions and derive estimates for received signal strengths at different localities. Such methods require high computational power and may also lack real network conditions.

Mobile operators also make use of the base station KPI parameters for detecting vertical handovers. The parameters corresponding to 3G IRAT HO (Inter Radio Access Technology Handovers) give an indication of the tendency of the mobile associated with a certain base station to experience outbound handovers from 3G access thus aiding the coverage analysis. Such parameters, however, suffer two major limitations: (i) they include handovers associated with connected-mode signaling only, and (ii) they only give an indication of "outbound" handovers, without specifying the destination base station in case the mobile changes the point of connection as well. Thus, an analysis based on such parameters fails to consider the idle-mode handovers, limiting the sampling space. It also misses geographical precision, as the lack of knowledge regarding destination base stations avoids pinpointing cells where 3G access is not adequate.

A cost-effective coverage analysis, that provides an accurate and precise identification of regions with 3G service discontinuity can be implemented through making use of the vast quantities of customer device data continuously collected by the operator- often classified as "big data", as outlined in the sections that follow.

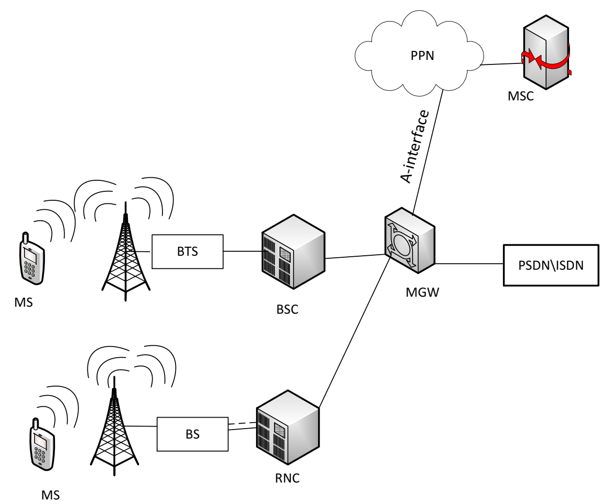


Fig. 1. Cellular network architecture and mirroring of A-interface in the network

A. *"Big- Data" and Hadoop Platform*

Technically, "Big data" refers to huge quantities of data that are so large, that it is practically not possible to process them using the readily available tools. Vast amounts of data that accumulates at companies as a result of their daily operations can be considered in this category. The analysis of big data is a central and important activity in a variety of fields such as finance, telecommunications, transportation and the like. Most high-technology companies such as Yahoo, Facebook, Google as well as many smaller companies are now processing terabyte-scale data processing on a regular basis.

The major technical challenge regarding big data is that it is not economically sound to store and process such data at its entirety, as the traditional analysis techniques are computationally expensive. To this end, advanced processing platforms specialized at handling big data, such as Oracle DB2, Microsoft PDW, EMC Greenplum, Teradata, Vertiva and Hadoop have recently been developed. Such platforms help lower the cost of storage, enhance computation and memory allocation.

Among the available platforms, Hadoop stands out as the most notable one as it is an open source solution. It is made up of a storage module, namely HDFS (Hadoop Distributed File System) and a computation module, namely MapReduce. Whereas HDFS can have centralized or distributed implementations, MapReduce inherently has a distributed structure that enables it to execute jobs in parallel on multiple nodes.

B. *Operator Data and Usage for Coverage Analysis*

Mobile operators typically are experiencing vast quantities of data that are being generated both by customer mobile devices (smartphones, laptops, etc) as well as various network elements present in their core networks (like MSC (Mobile Switching Center), MGW (Media Gateway), HLR (Home Location Register), etc), as illustrated in Fig 1. User generated data also include several million CDRs essentially processed on a daily basis, and includes information regarding phone calls, text messages, web site visits, online mobile banking transactions. Traditionally, among such huge quantity of data, only those related to billing, charging and similar customer relations functions are utilized by the operator. Data originating from network elements is often not considered in efforts to understand the user experience.

Such data can in fact be used to track inter-technology handovers. Number of such handovers in different localities are instrumental in identifying regions with interruptible 3G services. Towards this end, connected mode data inherent in the inter-node signalling in the form of IRAT HO (inter radio acess technology handover) indicators as well as the idle-mode data that corresponds to time intervals in which the mobile is silent can be utilized. An important source of such data in the operator network is the interface between the BSC and MSC nodes (also referred to as the "A-interface"), as also indicated in Fig. 1. The BSSAP messages that go through this interface provide the data needed to identify intertechnology handovers.

The BSSAP messages are mobile device based and include numerous indicators related to resource management, hand over control and mobility management. Among these numerous indicators, the four fields that are relevant in identifying inter-technology handovers are the location update fields namely "LAC serving", "Last LAC target", "BSSAP procedure" and "Last Cell ID" . LAC (location area code) serving field corresponds to the previous location area information of a mobile before location update and Last LAC/Last Cell ID target are the location area/cell information the mobile is currently attached to when the location update messages are sent. It is important to note that the location area information inherent in the LAC serving and Last LAC target fields also indicate the radio access technologies utilized by the mobile in the considered location areas.

The field value "BSSAP procedure = Update location" is indicative of handovers by users who reselect cells due to poor or lack of 3G coverage. Switch on or off actions, or 3GPP protocol enforced periodic location updates and mobility based location updates when a mobile moves from one location area into another in idle mod are also potential triggers. Thus, both idle mode and dynamic mode dynamics are captured in location updates. A joint analysis which takes into account previous and current location information reveals inter-technology handovers. The number of such handovers is indicative of 3G service discontinuity in the considered geographical location. It is also important to note that in such analysis, the cell id to which the mobile switches to 2G access can precisely be identified, providing the geographical precision that is missing in base station KPI parameters.

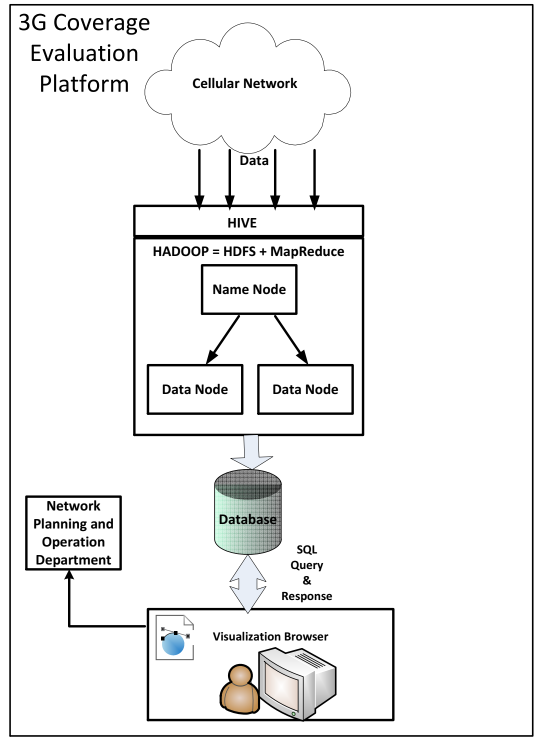


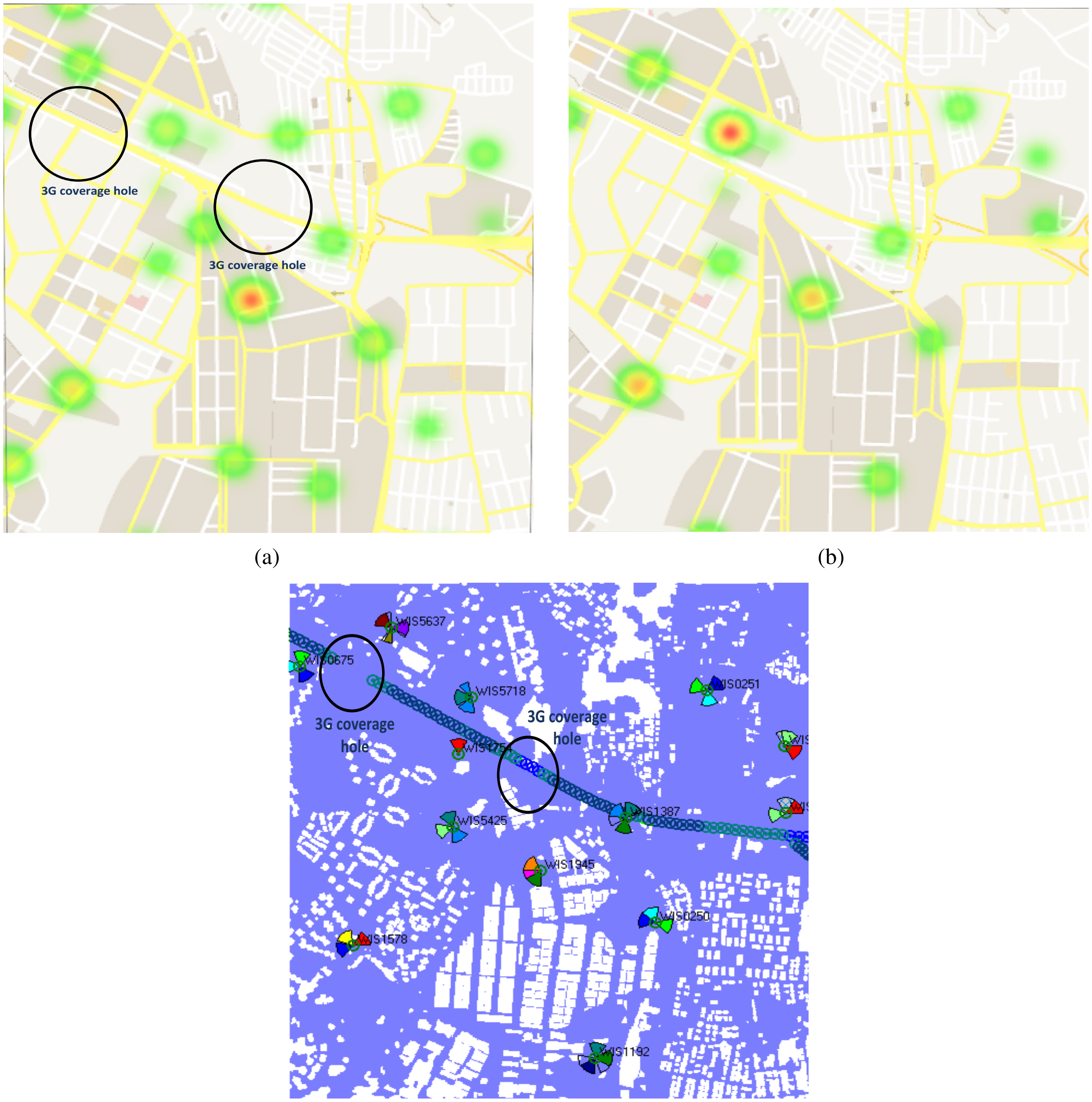
Fig. 2. Proposed 3G coverage evaluation methodology.

C. *Coverage Evaluation Methodology Based on Big Data*

A high level description of the proposed methodology that makes use of the A-interphase data is presented in Fig. 2. The proposed process for near-real time coverage performance evaluation methodology contains the following steps:

1) Data flows from cellular network to Hadoop distributed data warehouse (Hive), and is loaded into HDFS,
2) The data is queried with Hive in order to count all handovers (such as from 3G to 2G). Hive transforms the query to MapReduce functions implicitly,
3) Results are transferred to MySQL database
4) Relevant information from database is queried using SQL for only handovers from 3G to 2G. (This query can also be merged with the second step, but we decided to get all the handovers in order to not constraint ourself on specific IRAT HOs)
5) Coverage results are monitored on the web based GUI over google maps for the desired IRAT HO.

In step 1, the four relevant fields of LAC serving, Last LAC target, BSSAP procedure and Last Cell ID are loaded onto the Hadoop platform. In step 2, only those entries for



(c)

Fig. 3. (a) Google map illustration for 3G coverage hole identification using the proposed platform. (b) Base station IRAT HO KPI heatmap. (c) Drive test results on the same area.

which there are location updates ("BSSAP procedure = Update location") are determined, using the query in Table I. In step 4, the location updates are jointly considered with previous and current location information to identify handovers from 3G using the query in Table II. The Last Cell ID target field pinpoints the geographical locality of the handover. Note that high level query languages such as Pig Latin, Hive or Jaql can be used to express complex analysis tasks. Using extensions, queries are converted into MapReduce jobs. In the scope of this work, Hive was used for querying the network data in parallel.

III. Experimental validation

The accuracy and precision of the proposed mechanism was tested in Avea's network. Towards this purpose, a prototype was implemented along the lines outlined in the previous section. A data processing platform was implemented through Hadoop on three nodes, with computations powers corresponding to INTEL $2 * 1$ CORE CPU, 8 GB RAM, 500 GB hard disk.

The platform was fed with AVEA's A-interface data, which receives up to 1 Million messages a day that are generated and transmitted from different network elements in different parts of the network. The experimental demo was run on a sample set of data, that corresponds to 19-th to 20-th November 2012

TABLE I
HIVE QUERY

```
SELECT  lastcellidtarget,
        lacserving,
        lastlactarget,
        count(1)
FROM network cdr
WHERE bssapprocedure = ”Update location”
AND lacserving <> lastlactarget
GROUP BY lastcellidtarget,
         lacserving,
         lastlactarget,
```

TABLE II
SQL QUERY

```
SELECT  ns.latitude, AS lat
        ns.longitute, AS lon,
        sum(weight) AS weight,
FROM network cdr c
INNER JOIN networkstructure ns ON
ns.city = substr(c.lastlactarget,2,2)
AND ns.cellId = c.cellid
WHERE substr(c.lacserving,1,1) = ’4’,
AND c.lacserving <> ”65534”,
GROUP BY ns.latitude,,
         ns.longitude,
```

collected over the entirety of Turkey. The considered data is more than 64 GB in size and has a total of over 2 million rows of signalling and communicating information.

Representative results of the coverage analysis are presented in Fig. 3, which shows the web-based Graphical User Interface (GUI) of the prototype that allows monitoring the IRAT HOs. Fig. 3 (a) provides a screen shot of 3G coverage analysis map for a specific site IS1945 in Turkey, while Fig. 3(b) shows the corresponding area as it is analyzed through the base station based KPI. Fig. 3 (c) provides the results of the roadtest as benchmark. The hand over counts specify the color of the discontinuity spot. As can clearly be seen from the figures, the proposed method yields results that are in coherence with the drive tests, and are thus accurate. The comparison between the proposed method and the KPI based approach reveals that the proposed method provides greater precisions, identifying areas with service discontinuity that are not discovered by the latter method. These representative results demonstrate that inclusion of idle mode signaling data and target cell id info increases the quality of the coverage analysis.

IV. Conclusions and Future Work

In this paper, we have proposed a 3G coverage analysis method that identifies 3G service discontinuity areas by processing BSSAP signaling data through the Hadoop based big data processing methods. The proposed method identifies the location updates that correspond to inter-technology handovers and uses the target Cell ID provided to precisely identify geographic locality where there are 3G service discontinuities. The experimental results suggest that the proposed mechanism produces accurate results, and provide precision gains as opposed to traditionally employed methods. The approach displayed in this paper is a nice example demonstrating the use of big data concepts in increasing the efficiency of network operations in mobile operators. As future work, we aim to enhance the proposed method to enable near-real time analysis.

V. Acknowledgement

We thank Habib Korkmaz for providing evaluation results.

References

[1] C. Deng, L. Qian, M. Xu, Y. Du, Z. Luo, and S. Sun, “Federated cloud-based big data platform in telecommunications,” in *Proceedings of the 2012 workshop on Cloud services, federation, and the 8th open cirrus summit*, FederatedClouds ’12, (New York, NY, USA), pp. 44–48, ACM, 2012.

[2] Y. Dong, Q. Ke, Y. Cai, B. Wu, and B. Wang, “Teledata: data mining, social network analysis and statistics analysis system based on cloud computing in telecommunication industry,” in *Proceedings of the third international workshop on Cloud data management*, CloudDB ’11, (New York, NY, USA), pp. 41–48, ACM, 2011.

[3] W. Indyk, T. Kajdanowicz, and P. Kazienko, “Cooperative decision making algorithm for large networks using mapreduce programming model,” in *Cooperative Design, Visualization, and Engineering* (Y. Luo, ed.), vol. 7467 of *Lecture Notes in Computer Science*, pp. 53–56, Springer Berlin Heidelberg, 2012.

[4] H.-D. J. Jeong, W. Hyun, J. Lim, and I. You, “Anomaly teletraffic intrusion detection systems on hadoop-based platforms: A survey of some problems and solutions,” in *Network-Based Information Systems (NBiS), 2012 15th International Conference on*, pp. 766 –770, sept. 2012.

[5] J. Magnusson and T. Kvernvik, “Subscriber classification within telecom networks utilizing big data technologies and machine learning,” in *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, BigMine ’12, (New York, NY, USA), pp. 77–84, ACM, 2012.

[6] Z. Qiu, Z. wen Lin, and Y. Ma, “Research of hadoop-based data flow management system,” *The Journal of China Universities of Posts and Telecommunications*, vol. 18, Supplement 2, no. 0, pp. 164 – 168, 2011.

[7] W. Indyk, T. Kajdanowicz, P. Kazienko, and S. Plamowski, “Mapreduce approach to collective classification for networks,” in *Artificial Intelligence and Soft Computing* (L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, eds.), vol. 7267 of *Lecture Notes in Computer Science*, pp. 656–663, Springer Berlin Heidelberg, 2012.

[8] *Apache Hadoop, http://hadoop.apache.org/.*

[9] W. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire, “Minimization of drive tests solution in 3gpp,” *IEEE Communications Magazine*, vol. 50, pp. 28 – 36, June 2012.

[10] J. Puttonen, J. Turkka, O. Alanen, and J. Kurjenniemi, “Coverage optimization for minimization of drive tests in lte with extended rlf reporting,” in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, pp. 1764–1768, IEEE, 2010.

[11] *3GPP TS 37.320 v10.0.0, Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2.*

[12] B. Sayrac, J. Riihijarvi, P. Mahonen, S. Jemaa, E. Moulines, and S. Grimoud, “Improving coverage estimation for cellular networks with spatial bayesian prediction based on measurements,,” *ACM CellNet 2012 in conj. with ACM SIGCOMM 2012*, August 2012.