A Solution for Multipurpose Location Analysis over Bulk Real Time Data in Mobile Networks

Orçun Uzun
Network Services Department
Turkcell Technology Research and Development
Gebze, Turkey
Email: orcun.uzun@turkcellteknoloji.com.tr

Abstract—The goal of this paper is to introduce a new approach for location based analysis over big data volumes in a mobile network environment. Some use cases and possible functionalities for time and location based subscriber data and subscriber density have been discussed. An implementation for the approach is given. A simulation result is given for the tests that have been run over the implemented model. Many new business cases have been discussed.

Index Terms—location, big data, mobile network, subscriber data, location analytics, subscriber density

I. INTRODUCTION

Location has become a major information element for any corporate, organization or any individual. For anyone, it is great importance to have the location information whenever possible. For corporates and organizations, it would be possible to propose location based advertisements for a bulk customer community or for a non-profit organization it would be possible to check the location information whenever necessary or in case of emergency.

As mobile operators do have huge real time data within their network, it would be possible to give many important solutions and approaches. The most important data type for a mobile operator is the location information. Because voice, data or text services are some kind of legacy services so that the major differentiation will begin with what you can do using location data. Location Based Services (LBS) [1]-[10] provide very valuable capabilities for the mobile business. In many aspects, mobile business analytics depend on location data and their real time implementation [2], [3], [6].

The other important thing for a mobile network is processing the continuous bulk network data. This corresponds to dealing with the Big Data [4], [5] concept, but with a much more real time manner. Working with bulk real time data is a challenging task in mobile networks as any location change, data or voice signaling generates real time valuable information. For a large network that has over 40 million users, the management of real time location and event based data and generation of valuable high speed output becomes harder and harder.

From the mobile operator point of view, it is possible to provide real time bulk location data management capabilities such as user location determination, finding the number or the exact identities of users in a given location, time based location determination, movement and density analysis. These capabilities would make a great difference [7] for any organization that would be able to provide their customers or users.

Before this study, there was no such a solution design that is capable of analyzing the whole network subscriber activity within a mobile network and producing valuable real time and near real time location, movement and subscriber density information. There are several solutions that is capable of finding and analyzing the location of subscribers in a mobile network, but many of them triggers active queries to core network nodes to that increase the signaling traffic both in core network and radio network. For the analysis of a group or all of subscribers in a very near-real time, the network resources vanish very quickly, further; may lead to failures and decrease in Quality of Service.

With this study, a practical and feasible solution has been researched. The aim was to design such a solution that will provide many kind of location analysis using the whole real time signaling traffic within the network without disturbing any active network node. The research topic is to provide an affordable solution that will be able to analyze the subscriber movement, activity and patterns continuously within a large mobile network. Currently there is not any practical solution deployed in a mobile network that provides many location based data analysis over the whole network without introducing additional traffic load. The trials have been done with real data not synthetic. So that it had been possible to design a real-world solution that will lead to many new location based services, business opportunities and future challenges. Study summarizes good examples for real-world mobile operator experiences.

In this paper a practical approach for a multipurpose location analysis in a mobile network is presented. Many real world applications and use cases and the feasibility have been discussed. The next section shows the main characteristics and an overview of mobile networks. The third section gives the general overview of the approach and the uses cases that can be handled. The fourth section gives information about an implementation for the
approach. The fifth section gives simulation results and the final section sums up the studies.

II. MOBILE NETWORKS

Fig. 1 shows the general architecture [9], [10] of a basic mobile network.

Mobile Stations (MS) communicates with the mobile network via Base Station Subsystems (BSS) over the radio interface. Mobile network can be a hybrid of Circuit Switched (CS) and Packet Switched (PS) Core Network. BSS (Case Station Subsystem) communicates with Mobile Switching Center (MSC) and Serving GPRS Support Node (SGSN).

BSS section is composed of Base Transceiver Stations (BTS) that communicates directly with MS and Base Station Controllers (BSC) for 2G networks or Radio Network Controllers (RNC) for 3G networks for the entire control of group of BTSs and the signaling among radio part and core network.

In case of network requirements (voice or text) BSS communicates with MSC, for the data communications cases BSS communicates with SGSN. Visitor Location Center (VLR) is the real time temporary location center where the service and profile details of the subscriber are stored for the location where subscriber is served.

Home Location Center (HLR) is the central subscriber database that holds the service and profile details of subscribers. HLR communicates with VLR in case of CS location changes or with SGSN in case of PS location changes. Gateway GPRS Support Node (GGSN) is the gateway to the external packet network such as Internet. Gateway MSC (GMSC) is the gateway of CS part of the network to the outer CS mobile networks.

Each node in a mobile network has a standard signaling interface to the nodes that it needs to communicate. During any activity on a mobile network many important data is flooding over the nodes. Some of them can be stated as:

- Subscriber Identity (SI). In a mobile network it is International Mobile Subscriber Identity (IMSI) that uniquely identifies a subscriber.
- Activity Class (AC) that can be data, voice, text, location change etc.
- Activity Time (AT) that can be considered as the time that the activity has occurred.
- Location Information (LI) can be considered as the location of the cell that MS has been served, where the many cells can be grouped in a single BSS. Any base station may have more than one cell depending on the antenna types deployed. So that LI should be considered as the location of the base station that the MS is connected.
- User Equipment (UE) that can be considered as the brand and model name of the mobile equipment that subscriber uses.

III. OVERVIEW OF THE APPROACH AND USE CASES

Real time and near real-time individual and bulk location data can be very valuable source for many applications, tools, algorithms and products.

A. Real Time Location Information Gathering

Without disturbing mobile network resources via generating core network location queries to MSCs, SGSNs, HLRs, it is possible to query the LI of the subscriber in real time via storing the LI of the each subscriber in a cache with the participation of the other important parameters.

The LI of many subscribers can be queried in parallel so that many user applications that need a rapid response to bulk location queries can be handled easily. Location based bulk advertisement notifications can be a good example usage of this functionality.

This functionality has many use cases in a mobile network:

- Location based bulk advertisement notifications: Any location based notification can be sent to the subscriber after querying and analyzing the LI of the subscriber.
- Paging requirement analysis: In some kind of fresh location providing, MSC sends paging signaling to BSSs so that the subscriber is paged among all the cells that MSC cover. Paging introduces additional radio and core network signaling so that if possible and not required it should be prevented. If you know the age of the paging signals. So with this approach it is possible to retrieve the age of location via analyzing the last activity timestamp.

B. Data Extraction for a Given Location

In some cases the data at a given location may be required. For mobile operators the most valuable data in a given location is the subscriber. There are some other filtering data such as segment, age, UE type. Basically, a user application or a service may require the list of SIs in a given LI so that may provide special services for the subscribers served at that LI.

There may be several use cases that can fit to this functionality:
• **Emergency cases**: For an earthquake case the SIs in a given LI can be provided to a requester in a very high speed so that the required action can be taken instantly. As it is possible to store the SI and LI with the AT, it is possible to filter the list of SIs in given locations within the current time and AT. So that it is possible to find the identity of the subscribers existing in disaster area while the earthquake had occured.

• **Segment and user equipment based informative notifications and analysis**: Some user applications and services may require the SIs that uses a given UE and matches a given segment. So that mobile operators may easily gather the required information from the extracted cache and send some UE specific links or URLs to provide UE specific content.

C. **Determination of a Home Zone of a Subscriber**

Many of the operators provide Home Zone services for their subscribers. Some of them are the low cost voice, data or text usage in a determined LI. In order to give such kind of service the home LI of the subscriber should be provided by the subscriber via a voice call or SMS text so that the LI can be retrieved from the generated signaling in core network.

Via analyzing the LI of subscribers for a determined duration the analyzed LI may be considered as a Home Zone and then the subscriber can be informed to confirm that the location is the home zone of that subscriber.

D. **Movement Analysis**

Mobile network operators may provide movement analysis capabilities to transportation, airline or local service companies. It is possible to give any requester service the movement of a subscriber precisely. For example, the list of SIs that flies from city A to city B within a determined time interval. Using this information a potential airline company may make more realistic estimations during flight pre-planning.

E. **Data Density Analysis**

The other important use case for the location analysis over big volume data is the data density analysis. It is really extremely valuable for a mobile network operator and their customers to have the density and distribution of subscribers in a whole network.

This functionality has a very potential usage in real world applications. Think of a bank that tries to find the best locations for the Automated Teller Machines (ATM) that they need to introduce throughout the whole country or in a single city or in a small town. Then the density of bank’s subscribers within different time ranges would give very smart and valuable information in order to decide on the true location of the devices.

Any distribution information can be provided for a determined time range for the whole covered country. This information can be provided very rapidly to the requester service as it is possible to store the number of subscribers served within a given LI for a given time period.

It is always possible to find the total hit counts for a determined location range in a determined time range. It is also possible to give the total unique SI counts within the requested location.

Adding segment, UE model, age, gender like information to the density data provides very valuable social analysis for a mobile operator.

So that it is totally possible to show the exact density of group of or whole subscribers on a map on the whole country, city or district bases.

IV. **AN IMPLEMENTATION FOR THE APPROACH**

The use cases and approaches given in the previous section can be implemented with a distributed model.

In order to be able to gather the real time data flooding through the whole network, one of the best ways is to put a monitoring probe to each node and get the signaling of each one without disturbing the content. For the implementation, passive network monitoring probes has been inserted among main network nodes to be able to monitor the whole signaling occurred.

The monitored interface signaling can be processed in such a way that only necessary parameters are parsed. These parameters can be written into periodically generated files that will be analyzed further by location intelligence platforms.

The periodically generated files may be fed into location intelligence process. The location intelligence process is a high speed data processing flexible algorithm that is capable of processing the whole mobile network data in real time and make analysis depending on the queries.

As a real-world implementation, TURKCELL network has been used. With about 40 Millions of subscribers and daily 1 Million inbound roamers, this mobile operator maintains very huge volume of radio and core network signaling. The interfaces between the all MSCs and BSCs, SGSNs and BSCs, SGSNs and RNCs, SGSNs and BSCs had been monitored and processed using the deployed probes. At least 100000 distinct valuable access network and core network events (voice, SMS, mobile data/internet usage, location change, hand over, paging etc.) occurs with the whole monitored network. This makes more than 8 billion distinct events in a single day.

A. **Solution Overview**

As depicted in Fig. 2 the monitoring platform monitors the whole signaling among the man network nodes. Monitoring platform processes the signaling in a real time manner and produces large files that includes all the necessary information (SI, AT, AC, LI, UE, hand over, subscriber profile, web site visited etc.) periodically. The monitoring platform feeds the location intelligence platform with these files. The location intelligence platform reads the periodically generated raw network data files whenever they are available on the platform and stores in its real-time memory cache.

Within the memory all the location to subscriber list data mapping and subscriber to location information mapping may be stored. This memory can be checked
from a socket interface for real time location and subscriber data queries. The subscriber data list within a requested LI can be given from the cached memory data to the outer world; the requesting applications. So that analyzer understands the type of query and consults the memory cache for the required data and then passes the result to the requester. It is always possible to give any very near real-time location and activity data filtered by subscriber profile (age, income, work type etc.), UE, AT. The list of subscribers that resides in queried locations can be provided in real-time value ranges as well.

Analytics module is a run-per-request module where runs location analytics algorithms. For any request, module searches through the periodically generated snapshots, and runs its analysis over the data gathered from the snapshot files. The snapshot files generally contain distinct subscriber data with location information. There may be several types of snapshot files. One of them may be to hold location information to subscriber count mapping. That is, the count of subscribers or roamers within a given location in real-time. The other one may be to generate files where the location information is written in the file name and subscriber data is the content. The snapshot files are coordinated in an indexed manner so as to make the file search easy and quick.

So that whenever a density request is received, the analytics module directly analyses the requested time range and determines the folders and files that it is going to be opened, read and processed. The request may contain a list of LIs or an indication of whole network. Also the request may contain a subscriber list and then the analysis will be run only over the requested list.

Following density analysis can be done in a very high speed:

- **SI list based queries for the whole network:** In this scenario system will read all the snapshot files that store distinct subscriber information for the requested time range and filter by the provided SI list. SI list is initially written into a memory map so as to speed up filtering. For a determined group of read record the worker queues are fed. The worker queues can be extended with a simple configuration. The worker queue handlers processes the group data in parallel so that the result can be provided in a short while even for a large number of subscriber and processed record.

- **LI list based queries:** In this scenario system will read the location to subscriber data mapping or location to subscriber count mapping files. These files are short files when compared to distinct snapshot files so that it will be extremely shorter than the former one. An example for this kind of case will be to determine the subscriber density distribution for a city in the country.

The distribution of density values is an important task as it would be possible to distribute the density among the smallest possible site. So that having a database that holds the radio coverage distribution of a LI among several sites and grids, it would be possible to give approximate density information per a site or town or a street as well. In order to give very detailed density information system is mapped to a database that holds coverage distribution of LIs among sites. So that using a Graphical User Interface, it is possible to highlight the densities on a map and give the detailed one when zoomed.

### B. Implementation and Solution Overview

The implementation includes following hardware components

- A monitoring node that listens all the probes connected to it and generates configurable network data to periodically generated files.
- A workstation that has a 16 Core CPU with 2.67 GHz speed and 64 GB RAM to implement real time location query
- Monitoring node sends signalling information files every 15 seconds. Actually the period is configurable but 15 seconds have been chosen while testing.
- Each file contains at least 5 Million rows.
- System reads transferred files whenever they are available and writes into memory cache. The file...
reader modules has a multi-threaded architecture so that many files can be processed parallelly.

- Memory cache has been designed in such a way that support location to data and data to location mappings.
- Current mobile network has around 40 Million online subscribers so that UE, SI, AT and LI based mappings are all implemented for this subscriber volume. The data processed is the real time network data.
- For any query, this memory cache is checked and the result is returned instantly.
- In parallel, for a determined period, the snapshot files are generated using the cached memory data. This snapshot files are also used for density and location analytics.
- A workstation that has a 16 Core CPU with 2.67 GHz speed and 64 GB RAM to implement density analysis.
- System uses the snapshot files that are generated by the real time enquiry system.
- Each distinct subscriber data storing snapshot file stores about 40 Million rows.
- Each location to subscriber data count snapshot file stores about 110.000 rows.
- System accepts one query at a time, no parallel requests are allowed.
- For each query system goes into indexed file structure and finds the files to read.
- After reading the files, it sends the read data to worker queues.
- The density report is generated at the worker handlers using the coverage data available on the database that holds the coverage mapping.
- System DB and Map Info DB are standard MySQL databases.
- All the file processors, caching, request handlers, report processors, file generators, worker queues are built using C++ language on a Linux Operating system.
- Each workstation has about 2 Terabytes of storage.

V. SIMULATION RESULTS

There are approximately 40 Million distinct subscribers in the network and the subscriber data consists of SI, LI, UE, AT, AC and other subscriber segment data such as age, gender, and work title.

For every 15 seconds network interface data from 10 different interfaces are fed into system. For every 15 minutes following snapshot files are generated by the system:

- Distinct subscriber data (Approximately 40 Million rows)
- Files containing LI and subscriber list existing in that LI (Approximately 110,000 rows)

For density analysis simulation snapshot files are indexed in such a way that all the files with 15 minutes period is written into folders indexed from 1 to 7 representing days and from 0 to 23 representing hours.

Four different simulation tests have been run. CPU values are greater than %100 as the server has a multi-core CPU, so that %300 CPU usage means 3 CPU cores are being used.

1) Subscriber location query: The LI of subscriber are queried with a 4800 request per seconds traffic where periodic snapshot files are produced in parallel. The memory usage is about %80 and the CPU usage is about %950 where the workstation has 16 cores.

2) Subscriber data list query: The subscriber data list within given LIs are queried. There was 2500 request per seconds traffic where periodic snapshot files are produced in parallel. The memory usage is about %84 and the CPU usage is about %1100 where the workstation has 16 cores.

3) Density request for the whole network: The density and distribution of subscribers within the whole network is tested from Monday to Sunday from 00 to 23. The density analysis resulted in 102 minutes where the average CPU usage was %600 and memory usage was %85.

4) Density request for 20 million subscribers: The density and distribution of 20 Million subscribers is tested from Monday to Sunday from 00 to 23. The density analysis resulted in 330 minutes where the average CPU usage was %750 and memory usage was %80.

These implementations clearly shows that even with a single middle range performance server, it is possible to make location based analysis over the subscriber volume of a large operator in a very near real time and real time manner. The tests have been run among the real network data and the test contents are the ones that may be seen in a real-world mobile network.

The solution is designed for big real time location based data analysis for mobile operators. The summarized data within the real time cache and snapshot files are for 40 Million subscribers, but by enlarging and enriching the data with other external databases and mappings; and by storing the data trend for a long duration, the result will be a very huge data to be analyzed. For the cases where more parallel transactions or more diverse data is needed; so that a higher CPU, memory and storage capacity is required, systems that have data ware house capabilities.

VI. CONCLUSIONS

An efficient way of implementation and use cases have presented for multipurpose location analysis has been presented in this paper. Location and data analysis within a mobile network over a big data volume is challenging task. By implementing a solution over a big
sized mobile network, it has been observed that using big data volumes it is possible to provide very valuable information. Also it has been observed in simulations that with large and complex big data systems it is really possible to have instant any content tagged location and density analysis in a big volume mobile network.

There are many other use cases where the solution suits, but only some of them has been introduced. For example, with this solution, it is quite possible to find the failures within the radio network by analyzing the inbound roamers that have been lost by the network and being served by other operator in the country.

ACKNOWLEDGMENT

This work has been performed as a research activity as a part of Location Based Intelligence studies by Turkcell Technology Research and Development. The author would like to thank all the people in Turkcell Technology for their support.

REFERENCES

[10] 3GPP TS 22.071 v 11.0.0, Location Services (LCS).

Orçun Uzun was born in Adana, Turkey in 1981, November 30th. In 2003 he has earned Bachelor of Science degree on Electrical and Electronics Engineering from Çukurova University, Institute of Science, Adana, Turkey. In 2006, he has earned the Master of Science degree on Electrical and Electronics Engineering from Boğaziçi University, Institute of Science, Istanbul, Turkey. Currently he is working on modern mobile and cellular telecommunication networks, product design and development. Location based systems, core network products, next generation telecommunication systems and algorithms are the basic field of studies. He had completed the military task in 2007 in Mardin, Turkey. He has started his work experience in Erel Technology as a Project Engineer in 2004. In 2007, he has started Nortel Networks Netas as a Telecommunications Product Design Engineer and had worked on next generation systems and participated in design and development of many products and features. In 2010, he has joined Turkcell Technology, located in Gebze, which is the R&D part of Turkcell, the biggest mobile operator in Turkey, and one of the biggest mobile operators in Europe. He had worked in many projects, designed and developed many new products as a Product Design Engineer. From May 2012 to present, he is working as the Technical Product Manager, being responsible for the management of the design and analysis of the Mobile Network products of Turkcell Technology.

Mr. Uzun is a member of IAENG and IEEE.