

## THE CELLULAR NETWORK - AN EMERGING RESOURCE FOR CITY PLANNING

Mrs. Roohana Parabia, Dr. Paresh Virparia, Dr. Sanjay Buch

### ABSTRACT

Telecommunication companies generate a tremendous amount of data. These data include call detail data, which describes the calls that traverse the telecommunication networks, network data, which describes the state of the hardware and software components in the network, and customer data, which describes the telecommunication customers. There are several applications where telecommunication data mining can be used to uncover useful information buried within these data sets. The primary applications are, to identify telecommunication fraud, to improve marketing effectiveness, and to identify network faults. The new emerging telecom data mining applications are industry specific like bank marketing system, city planning system, census data management system, etc. In this paper we have presented a Telecommunication Data warehouse Model to discover knowledge regarding mobility patterns and traffic density in the cellular network for the prediction of road demand that can be helpful to the city planners to make effective decisions using cellular network usage data.

**Keywords:** Data Warehouse, Road Traffic Management, Telecommunications.

### 1. INTRODUCTION

Mobile phone datasets have become widely available in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, build trust, create markets and develop social interactions. Mobile phone data is also helping us understand complex processes such as the spread of information and viruses or transportation and the use of urban infrastructures. [1]

Researchers in academia, and increasingly within the mobile industry, are working with large databases showing where and when calls and texts are made and received to reveal commuting habits, how far people travel for public events, and even significant social trends.

With potential applications ranging from city planning to marketing, such studies could also provide a new source of revenue for the cell phone companies. "Because cell phones have become so ubiquitous, mining the data they generate can really revolutionize the study of human behavior," says Ramón Cáceres, a lead researcher at AT&T's research labs in Florham Park, NJ. Here we will study the use of mobile data for city planning. [2]

• **Privacy Issues:**

Data mining applications must always consider privacy issues. This is especially true in the telecommunications industry, since telecommunication companies maintain highly private information, such as whom each customer calls which is stored in form of Call Detail Records (CDRs). Most telecommunication companies utilize this information conscientiously and consequently privacy concerns have thus far been minimized. A more significant issue in the telecommunications industry relates to specific legal restrictions on how data may be used. Thus, we are trying to exploit a new method in a live GSM network for getting the traffic data (moving patterns) to build a recommendation system for city planning.

Our recommended system needs to gather the real time series of network traffic data from cellular networks for the purpose of analyzing moving patterns and population density in a given location area. It has no concern with the CDRs. This data is generated periodically by Base Transceiver Station (BTS) in the cellular network. The data is in the form of total number of Mobile Stations (MSs) present in a given BTS at a particular time instance. These are aggregate data, which doesn't reveal the identity of a user or any call initiated by it. Thus there is no concern of privacy issue. Certainly there are question and technical problems had to be solved. Example: How to get suitable location information? How often and how long to track the network usage data to get the best result while not increasing the load of the network unnecessarily? How to process large amount of collected data to get desired information in comprehensible format?

• **Why should we rely on mobile data for such kind of studies?**

Since almost everyone has a cell phone, the scale of the data is immense compared to other sources. Mobility patterns might, for example, be used for city planning, to provide accurate traffic forecasts for special events, cell network usage management and many such similar kind of applications.

• **How this information sharing can be helpful to telecom industry?**

Cell phone networks can have a "two-sided" business model by monetizing their data where they not only serve end users but also make money through relationships with other businesses. "This is new in the telco space".

In this paper we will first discuss regarding the current methodology used in city planning and its limitations. In second phase we will discuss regarding the new analytical applications (under research) built on mobile data (CDRs) to assist decision making for city planning and their limitations. The third phase describes the proposed Telecommunication Data warehouse along with its benefits which is useful for city planning and similar such applications.

## **2. HOW DO CITY PLANNERS WORK? (EXISTING SYSTEM)**

### **Manual Data Collection:**

City planning till today is done manually. Volunteers are appointed to do the survey of the traffic regions. Depending on their submitted information and the rules of the government generally the decisions are made.

### **Pitfalls:**

- It consumes lots of effort and is totally dependent on volunteers' feedback.
- It has no verification mechanism to check the correctness of their feedback.
- It is very expensive

## **3. ONGOING RESEARCH ON THE CITY PLANNING USING MOBILE DATA**

### **A. Call Detail Records (CDRs) are Collected as a Source**

The data set is a collection of call detail records, or CDRs--the standard feedstock of cell phone data mining. A CDR is generated for every voice or Short Message Service (SMS) connection. Among other things, it shows the origin and destination number, the type and duration of connection, and, most crucially, the unique ID of the cell tower a handset was connected to when a connection was made. [1][2]

We can know the location of a phone to within a mile radius at the time each CDR was generated, making it possible to determine the distance travelled from home by each cell phone every day. These details can be helpful to us to know how far people travel most of the days. Thus, city planners, who would usually have to resort to expensive and limited surveys to gather such information can

use this information to know where to build a new train, a station, a bridge or roads.

## **B. Problems faced by the researchers**

- **Security issue:** - Research in this area is typically focusing on aggregate information and not individuals. We can protect user privacy by removing private details (like names and numbers) from a CDR but still correlating locations and call timings with other databases could help identify individuals. Example, could infer the approximate home location of users by assuming it to be where a handset was most located between 10 p.m. and 7a.m. It is our responsibility for finding out how to trade off having useful data and protecting privacy.

- **Complexity of data retrieval:** - We can't use CDR directly as it contains information regarding individual event. It contains all the details regarding a call like call duration, call time, location information, caller identity, called user identity, last base station the mobile has entered while call was active, etc. We have to retrieve the required data from this database, clean it, transform it according to our needs and then store it inside the telecom data warehouse. Thus, extract, transform and load (ETL) process for this data warehouse is little complicated.

- **Not a reliable Data Source:** - Here we are using the time series data of call detail records (CDRs). It is not necessary that a mobile should initiate a call or send an SMS while he is in that cell region. For example, if there are 10,000 mobile users found in a cell region to be active and only 8,000 mobile users are initiating a call or sending a SMS then we will have details of only 8,000 users in a CDR database. So, instead of showing the population density as 10,000 it will reveal only 8,000. Thus, our predictions can be wrong, as our input sources are not reliable. Thus, to get the correct results it is assumed that each mobile user will either initiate a call or send a SMS, which is not true otherwise.

## **4. PROPOSED AYATEM**

Our system gathers the data from the cellular network and stores it in the format that can be helpful to us in future for firing managerial queries (city planning). And not only that many similar applications can be developed which can give us insight in the social trends of people. They store the time snapshots of traffic density in a given area. It generates a huge managerial database from which

telecom industries can earn money and the city planners can reduce the cost and effort of finding required information. [3][4]

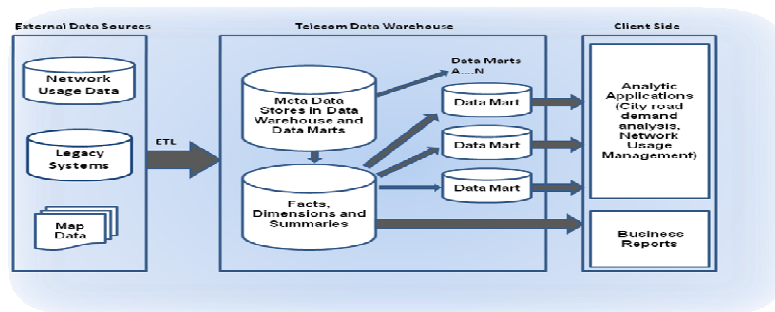


Fig. 1 Telecommunication Data Warehouse model

## A. External Data Sources

### 1) Network Usage Data collected as a source:

The data set is a collection of details generated by Common Control Channels (CCCH) and the Broadcast Control Channel (BCCH). Each active handset is connected to a unique Base station in the given location area. Each Base Station has the unique ID. Each mobile is required to regularly report its location at a set time interval using a periodic location update procedure. By getting the count of number of mobile stations entered in a given Base Station (BS) at a particular time we can get the traffic density at a given point of time at a particular location. Thus the time snap shots of the traffic density in a particular area can help us determine the requirement of a desired infrastructure (road or a railway).

### 2) Map Data:

Today, map data, also referred to as spatial data (data that reference location), is the fuel for location-based applications. Good data—precise, accurate, and reliable—is the essential foundation for useful information. With robust map data underpinning software applications, users can view maps and do much more:

- a. Plot a route to a destination—the fastest way, the most scenic way, avoiding freeways, and and so on.

- b. Sort categories, such as businesses, by time and distance from the user's location.
- c. Dynamically modify a route to avoid obstructions, such as heavy traffic.

Correct decisions often rely on complete, accurate, reliable, and up-to-date map data for navigation, traffic, point of interest (POI) lookup, social networking, entertainment and leisure, business applications like geomarketing, and industry vertical types of applications like telecom, mobile resource management, and the like.

Map data consists of topologically consistent vector or raster data that can be organized in various ways, such as in a relational database. This data is not only essential for calculating routes, avoiding traffic, and finding a nearby destination or POI, but is also useful for displaying or printing maps or for other types of geographic analysis.

### **3) Legacy Systems:**

These are older computerized systems which store Government rules and regulations data may be in Word or .pdf formats or any older softwares.

## **B. Telecom Data warehouse**

### **1) Facts and Dimensions:**

This is the component of the data warehouse that acts as the primary storage area for the data in the data warehouse. Typically this component is populated by data coming from the operational systems. The data structures in the fact tables (traffic density details) are typically generalized and the bulk of the Telecommunication Data warehouse would fit into this component. Dimensions are the tables that behave like scales to analyze factual data. (Ex, Time or Location). [4][5]

### **2) Summaries:**

This area contains and describes summaries and aggregations that are commonly and frequently used throughout the enterprise. These summary entities are typically (but not necessarily exclusively) populated from the Fact Data. Creating and maintaining such summaries in a data warehouse assists in ensuring a standardization of such summaries across the enterprise. These summary

entities are designed to store key measurements and indicators on a periodic (typically monthly or quarterly) basis.

### **3) Data warehouse Metadata:**

They are pieces of information stored in one or more special-purpose metadata repositories that include (a) information on the contents of the data warehouse, their location and their structure, (b) information on the processes that take place in the data warehouse back-stage, concerning the refreshment of the warehouse with clean, up-to-date, semantically and structurally reconciled data, (c) information on the implicit semantics of data (with respect to a common enterprise model), along with any other kind of data that aids the end-user exploit the information of the warehouse, (d) information on the infrastructure and physical characteristics of components and the sources of the data warehouse, and, (e) information including security, authentication, and usage statistics that aids the administrator tune the operation of the data warehouse as appropriate. [7]

### **4) Data Marts:**

A data mart (DM) is the access layer of the Data Warehouse (DW) environment that is used to get data out to the users. The DM is a subset of the DW, usually oriented to meet the specific demands of a particular group of knowledge users in terms of analysis, content, presentation and ease of use. They contain specific aggregations or summaries which is required for a specific purpose. Example, we can create a separate data mart for city planners and census detail analysers. [5][6]

## **C. Client Side**

### **1) Analytic Applications:**

Analytic Applications are a type of business application software, used to measure and improve the performance of business operations. More specifically, Analytic Applications are a type of Business Intelligence solution. It addresses issues such as:

- a)** Network Usage Management
- b)** Demand Analysis for roads.
- c)** Accurate Traffic forecast for future events.

## 2) Business Reports:

It is possible for Operators to attach a range of business reports to the Telecom DW. Such solutions range from spreadsheets and data warehouse reporting tools, to specific applications in the area of network resource management, data mining, city road traffic analysis, population density details, etc. In the case of spreadsheets or data warehousing reporting tools once the equivalent data mart structures have been created it is then possible to rapidly create extensive and sophisticated reports and charts that have been customized to the exact needs of the business users. Using summary reporting the Operator can 'slice and dice' the information to quickly identify trends and patterns (user mobility patterns). Summary reports typically query data marts.

## 5. BENEFITS OF THE TELECOM DATA WAREHOUSE

- **More Secured:** - As we propose to use network usage data there isn't any issue of security. We are using aggregate data of control channels which has no concern with the traffic channel data (speech or SMS) also called CDRs.
- **Easy Data Retrieval:** - The data retrieval procedures are simpler and faster. We need to scan the dynamic (Visitor Location Register) VLR of the GSM network to get the location information of the mobile users. The scanning process is done periodically depending upon set time window. The time snapshots of VLR data are stored in our Telecom Data Warehouse.
- **More reliable Data Source:** - Since almost everyone has a cell phone, the scale of the data is immense compared to other sources. Each mobile station is connected to a unique Base Station all the time. So if we acquire the time series data of number of mobile stations active in a given cell region we can easily find the traffic density and mobility patterns.

## 6. CONCLUSION

Our findings based on the study done by us for utility of the mobile data for the various purposes are as under:

- Mobile devices are the portable data-capturing terminals, which would collect the data based on its position at a given point of time in the Mobile Network.
- The information collected through mobile network usage data, has a greater potential to cultivate or transform for the purpose planning various city resource.
- The network usage data also directs us about peak and ebb timing.
- It is propose to have structured approach by using the Data Warehousing/Data Modeling methods for assisting city planners.

- Various Information security dimensions involved needs further study to identify the suitable framework.

The findings described above are being explored in detail. With the progress of research we anticipate more challenges and constraints.

The competitive and changing nature of the industry, combined with the fact that the industry generates enormous amounts of data, ensures that data warehousing will play an important role in the future of the telecommunications industry.

#### REFERENCES:

1. M. Cochinwala. Database Performance for Next Generation Telecommunications. Proceedings of the 17th International Conference on Data Engineering, 2001 IEEE.
2. Tom Simonite. Mobile Data: A Gold Mine for Telcos. Technology Review published by MIT, 27th May 2010.
3. Telecommunications Data Warehouse (TDW) General Information Manual. IBM Information Framework 2003.
4. Gary M. Weiss. Data Mining in the Telecommunications Industry. IGI Global, 2009.
5. S. Anahory and D. Murrey. Data Warehousing in the Real World. Pub. Pearson Education.
6. P. Ponniah. Data Warehousing Fundamentals. Pub. Wiley India Edition.
7. Panos Vassiliadis. Data Warehouse Metadata

#### AUTHORS' PROFILE:



**Mrs. Roohana Parabia** is a research student in Computer Science from Sardar Patel University, Vallabh Vidyanagar. Telecom Data Warehouse is her main research area. She has 5 years of experience in computer academics field. She has worked as a lecturer with prestigious institutes like CMC, Surat and ITC Brahmjyot, V. V. Nagar. Lastly she worked as Assistant Professor in M.Sc. (I.T.) Department, V.N.S.G.U., Surat.

**Dr. Paresh Virparia** is working as a Senior Associate Professor in the Department of Computer Science of Sardar Patel University, Vallabh Vidyanagar. He completed his MCA in 1989 from Sardar Patel University and Ph. D. in 2002 from Sardar Patel University. He is recognized Ph.D. guide in Computer Science at Sardar Patel University, V V Nagar and Kadi Vishva Vidyalaya, Gandhinagar. Currently, NINE students are doing their Ph. D. under the guidance of him. Two students have completed their M.Phil. (Comp.Sc.) under his supervision. His publications include 8 papers in International Journal, 6 papers in National Journals and 30



papers in national conferences/seminars. His research interests include the areas of Computer Simulation & Modeling, Networking and IT enabled services.



**Dr. Sanjay Buch** is IT Professional, With PhD in Computer Science from V.N.S.G.U. Data Warehousing and Data Quality are his main research areas. He is also certified with Project Management Professional (PMP) certified by the Project Management Institute, USA & ITIL V3 by OGC-UK. He has more than 26+ years of cross functional experience in multiple IT domains such as Petrochemical & Polyester Manufacturing, Mutual Fund & Investor Services, Capital Market, Public Policy Research, Retail and Application Management/Operations at Enterprise level. Currently working in Reliance Corporate IT Park, Mumbai as Assistant Vice President.