Design and Development of Geographically Distributed Algorithm for SMS Based Applications

By

Nizam Hussain

2011-NUST-MS PhD-IT-029

Supervisor

Dr. Hamid Mukhtar
Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of
Masters of Science in Information Technology (MS IT)
In
School of Electrical Engineering and Computer Science
National University of Sciences and Technology
Islamabad, Pakistan
(July, 2015)

Approval

It is certified that the contents and form of the thesis entitled **"Design and development of geographically distributed algorithm for SMS based applications**" submitted by **Nizam Hussain** have been found satisfactory for the requirement of the degree.

Advisor: ___Dr. Hamid Mukhtar_____

Signature: ____________________________

Date: ___________________

Committee Member1: __Dr. Hassaan Khaliq___

Signature____________________________

Date: ______________________________

Committee Member2: __Dr. Raihan Ur Rasool_

Signature ____________________________

Date: _______________________________

Committee Member3: Mr. Muhammad Qaisar Choudhary

Signature _______________

Date: __________________

Dedication

I dedicate this thesis work to my parents who from the start encouraged and supported me. Also to my younger brother **Mr. Shakil Ahmed** who is an endless source of motivation and guidance for me. He has supported, encouraged, and motivated me throughout my educational career. This thesis is also the result of his persistent efforts, support, motivations and sacrifices for me. His sacrifices have not gone unnoticed and I can never thank him for all he has done for me.

It is also dedicated to my friends and to my teachers with whom I have an exceptional and admirable relationship.

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by any other person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: __**Nizam Hussain**__

Signature: __________________

Acknowledgment

First and foremost praises be to Allah, the Almighty, on whom ultimately we depend for direction and guidance. It’s a great blessing from Almighty Allah that gives me the health and strength to do this research work. After this I would pay my sincere thanks to my parents, who have given me their unequivocal support throughout my life.

I wish to express my deep sense of gratitude to my supervisor **Dr. Hamid Mukhtar** for giving me a chance to work with him, in a research environment and polish my skills. Many thanks for his patience, effort and valuable guidance. I have been very fortunate to have him as my advisor.

I am paying my special gratitude to **Dr. Hassaan Khaliq**, **Dr. Raihan Ur Rasool**, and **Mr. Muhammad Qaisar Choudhary** for serving as my committee members and for giving me brilliant suggestions relevant to my thesis work.

A special word of thanks goes to **Dr. Raihan Ur Rasool** for his critique and valuable suggestions. I have throughout enjoyed my discussion with him and have learned a lot. I also would like to thank him for the effort that he has showed in communicating with me from a distance via email and Skype calls whenever required.

Finally, yet importantly, I would like to express my heartfelt thanks to my friends, colleagues and fellow students who contributed to the success of the project and the writing of this thesis directly or indirectly.

Table of Contents

Contents

Acronyms and Abbreviations

LBS.. Location-based services
SMS .... Short Message Service
MSC.. Mobile Switching Center
BTS.. Base Tranciever System
BSS .... Base Station System
CH .... Cluster Head
SCH .... Sub-cluster Head
CM... Cluster Member
DHCA.. Dynamic Hierarchical Clustering Algorithm
CID .... Cell Identity
LAC .... Location Area Code
GIS.. Geographical Information System
LAC .... Location Area Code
ME .... Mobile Equipment
MS .... Mobile Station
GSM .... Global System for Mobile Communication
PSTN .... Public Switch Telephony Network
LAC .... Location Area Code
SIM... Subscriber Identity Module
MCC .... Mobile Country Code
MNC .... Mobile Network Code
MANET.. Mobile Ad Hock Network

List of Tables

List of Figures

List of Algorithms

Abstract

With large advances in the wireless network technologies, internet, and mobile computing devices such as smart phones that enable users to access latest updated information ubiquitously from internet in an "anytime, anywhere, and any cost" style on their mobile phone devices. However, in this model user always require internet connectivity such as data or Wi-Fi connection to access updated information. But, the most common challenge is that data connection is typically more expensive, Wi-Fi connection is usually not available everywhere, and often it is not even free for all users.

To access updated information on mobile devices at anytime and anywhere regardless of network connectivity with cheapest prices or even free of cost, a good option can be to deliver latest information from web server to users through short message service (SMS). The main idea is to utilize short message service to deliver information due to its pervasive availability and worldwide support by various mobile devices and wireless cellular networks. The primary goal of our research is to design and develop a geographically distributed algorithm for SMS based applications that delivers the updated information contents from source (web server) to various mobile subscribers based on their current geographical location via SMS. So we have proposed a geographical location-based dynamic hierarchical clustering (DHC) algorithm for grouping large number of cellular subscribers and head node selection algorithm for selecting some of the user's mobile devices as network relay nodes. We have used users' interest with their current geographical vicinity to make clusters while considered device capacity for selecting relay nodes. The performance and efficiency of proposed algorithms is evaluated and validated using various parameters by simulating it on Google maps using JavaScript and Google maps API. At the end we have developed a prototype system. This prototype system validated our design and proposed algorithms, and also evaluated the performance of proposed algorithms.

Chapter 1
Introduction

This chapter gives the basic idea of the concepts involved in this research. It also presents the background and motivation for this study. Moreover, it provides the objectives, hypothesis, gives an idea of expected results, and methodology to get and evaluate the results. Finally, it presents the structure of this thesis document.

1.1. Location-based services

Location-based services (LBS) join wireless networking technology and geographic information systems (GIS) functionality and manipulate location information to support mobile subscribers [11]. Classic applications of LBS such as finding the nearest service of a certain type, finding neighboring mobile users or nearby friends, or receiving alerts messages based on vicinity to certain locations and consequences[1]. Such type of services uses information about the relative positions of the mobile user and the services of interest or of other nearby users. Some of the key service areas of LBS are follows:

- Information and navigation services.
- Tracking and monitoring services.
- Emergency and safety services.
- Entertainment services.
- Weather and News alert services.

LBS are services accessible through a mobile cell phone and always consider the mobile device's current geographical location information. These services can be categorized into two types: push and pull[2]. In a Push type of service, the subscriber receives information messages after some time interval from the service provider without any request, while the subscriber may have initially subscribed to the service at a prior time. In a Pull type service, the subscriber has to dynamically request for interested information.

As the business for smart mobile phones, and other cellular devices, continues growing, the need for new services emerges to fascinate the end users. One point that is being talked about all through the world is users current location based services[3]. In what capacity can a cell phone location be located and in which way can a service be developed to use this location information? Work has started to make a standard of how the position can be obtained from the existing network, in the meantime different methods have been described. Some these methods requires some additional hardware components to locate mobile devices in network. So this forces end users to spend more money in order to update their devices[4].

This work provides a better solution how the current location of users' mobile devices can be obtained from the existing cellular network by using SIM (Subscriber identity module) card. The system then use this location information to group mobile users in different geographical regions and deliver information contents to interested users based on their group membership.

1.2. Research Problem

Mobile subscribers in wireless cellular networks always require network connectivity (wireless access point or data connection) in order to accesses information contents from internet on their mobile devices [25][5]. Today data or Wi-Fi connection is the most common and widely used method to connect with internet anywhere and anytime. But the most common challenge is that data connection is typically more expensive; Wi-Fi connection is usually not available everywhere and often it is not free for each mobile subscriber specifically in rural areas[6].

Beside the above described problem, most of the existing Distributed publish-subscribe and location-based systems for information dissemination are composed with challenges of overload, performance degradation, and inadequate network scalability [7]. This is normally caused by an imbalanced load sharing in-between network communication devices of real applications and the vulnerability of links failure in underlying networks specifically in wirelessly linked mobile devices [8]. Network splitting and load distribution techniques can be used to deal with such problems. This leads to the dynamic load distribution in run time and delivering of information contents in a fully decentralized fashion, seeks to increase the network performance by grouping mobile devices by considering their current location in wireless cellular networks. It attempts to successfully achieve a more reliable and equal load distribution and to avoid the occurrence of an abundant load on centralized server or other intermediate communication devices in network.

1.3. Motivation

The use of short message service (SMS) is increasing with time throughout the world especially in Pakistan [11]. In last couple of years, trend of SMS is growing fast in Pakistan and it is estimated that SMS traffic will grow double in coming three years and volume of messages will double to approximately 10 trillion by 2015 [18][10]. Different Cellular companies in Pakistan offer low price SMS bundle packages [1] to their customers, they also give liberty of sending unlimited SMS to any network in Pakistan in a very cheap prices [23][24].

To tackle the above mentioned problems and to provide an efficient location-based information services to mobile subscribers, we propose a scalable, decentralized, and geographically distributed system which enables mobile subscribers to accesses latest information contents to their interest anywhere and anytime style on their mobile phones. Network splitting and load distribution techniques can be used to deal with such problems. This leads to the dynamic load distribution in run time and delivering of information contents in a fully decentralized fashion, seeks to increase the network performance by grouping mobile devices by considering their current location in wireless cellular networks. It attempts to successfully achieve a more reliable and equal load distribution and to avoid the occurrence of an abundant load on centralized server or other intermediate communication devices in network. The main idea is to utilize short message service to deliver information due to its pervasive availability and worldwide support by various mobile phones and wireless cellular networks[9].

1.4. Objective of this thesis

The primary objective of this thesis is to design and develop a location-based, scalable, and decentralized system which delivers updated information contents from a

[1] http://www.pta.gov.pk/index.php.

centralized web server to a group of subscribers based on their current location information via SMS. Other objectives of the proposed system are as follows:

- **Design Objective.** Design and development of geographically distributed algorithm to implement the proposed system in such a way that some of the subscribers mobile devices will be used as an intermediate processing nodes referred as network relays or cluster/sub-cluster heads. These intermediate processing nodes have some additional functionality that they are not only limited to receive messages to their interest but they also capable to multi cast message to their cluster members**.**
- **User Clustering.** Grouping mobile subscribers to make clusters according to their interest with their current geographical location. Dynamically add or remove users from existing clusters; split, and merge clusters based on the size of subscribers in different geographical regions.
- **Message Distribution.** Develop a new mechanism for routing messages in wireless cellular networks using short message services. Messages are distributed through cluster and sub-cluster heads in order to minimize cost and communication load on server; and maximizes the efficiency and scalability of the network.
- **Prototype Implementation and Simulation.** Implement the proposed algorithms of user clustering and message distribution. Analyze these algorithms theoretically and through simulation. At the end of this thesis a prototype application is proposed. This prototype system validates our design and proposed algorithms, and also evaluates the performance of proposed algorithms.

1.5. Challenges

There are different challenges regarding location-based message distribution and most specifically related to wireless mobile network. Some of the main challenges are follows:

- **Users Mobility.** Mobile subscribers in wireless cellular network continuously change their location so that the system should be dynamic that reflects to changing in user's current location.

- **Diverse Mobile Platforms.** Currently, diverse mobile platforms are available in market such as viz. Symbian, iPhone, Android, Blackberry; Windows are top five most used mobile platforms nowadays in worldwide [25]. So there are lots of incompatible issues between different mobile platforms and required high development efforts and maintenance cost.
- **Cluster Size.** How to make clusters according to current geographical location information and how to fix a maximum threshold value for the size of a cluster? It is hard to adjust cluster size dynamically in different geographical regions.
- **Cluster Head Selection.** How to elect a node as cluster head. It is hard to assure the availability of cluster head in mobile network.
- **Messages or SMS limitation.** Typically short message service (SMS) support maximum size of 160 character text message [18] [19]. In the proposed system, all the communication between server and various subscribers through SMS. Therefore maximum number of SMS is needed by the system in order to timely deliver information contents. But we have the limited number of SMS from subscribers as they prescribed the size of free SMS at the time of service subscription.

1.6. Contribution

The contribution of this research is to design and development of a geographically distributed algorithm for cellular networks that supports geo-messaging. The outcomes of this thesis are following:

1. **Scientific Contribution.** We introduced a new dynamic hierarchical clustering and cluster head selection algorithm for large number of mobile subscribers in wireless cellular networks. Cluster formation process involves selection of a mobile device as cluster head. Each cluster consists of at most one cluster head (CH) and other non-cluster head nodes (cluster members) that form a cluster dynamically. In our proposed algorithm we feat the formation of mobile devices into clusters in order to reduce the communication load encountered in centralized web server and the related network overhead.

2. **Prototype of the System.** At the end of this thesis a prototype application is proposed. This prototype system validates our design and proposed algorithms, and also evaluates the performance of proposed algorithms.
3. **Geo-Messaging.** The outcome of this research is the distribution of information contents across the mobile subscribers according to their current geographical location through cluster and sub-cluster heads using SMS.
4. **Client Application.** The most important task of this thesis was to build the client application for Android platform which retrieves subscriber's location information from network and send to a centralized web server. The client application also provides a user interface through which subscribers set their preferences at initial service subscription. The application also receives stores, display, and forward messages to server and other subscribers. In addition, application also stores clustering related information on mobile devices, later it is used to deliver messages to other subscribers in the network.
5. **Dataset.** We created a dataset having a list of approximately 6,000 GSM network cell identities (CIDs) with their geographical coordinates and location area code (LAC) throughout Pakistan. The dataset consist of all five GSM network operators which providing cellular communication services in Pakistan including Telenor, Zong, Warid, Mobilink, and ufone. This dataset can be made available to other researchers interested in similar problem.

1.7. Applications

This prototype of the system for dissemination information can be used to deliver more personalized information contents to various mobile subscribers on their cell phones anytime, anywhere, and any cost style. Later, it can be used for:

- Diverse Location-Based services such as News alerts, Weather updates, and Traffic reports, etc.
- Location-Based Advertisements.
- Recommendation systems.
- Emergency and safety services.
- Location Based Gaming, etc.

1.8. Structure of this thesis

After the above introduction, chapter 2 will discuss related work done so for in the field specifically in wireless cellular networks. We will further discuss the state of art of the hierarchical clustering algorithms for grouping similar objects based on their current geographical location. In chapter 3, we will discuss system design, architecture, implementation, and brief introduction of its sub components. Chapter 4 gives detail discussion regarding the datasets, simulation, experiments, and results of experiments. Finally, the thesis will end with concluding remarks and some future works can be done in chapter 5.

Chapter 2
Literature Survey

The following paragraphs provide literature review related to the location-based services, GSM communication network, and short message service (SMS) and its applications in location based systems.

By using short messaging service[2], users can send messages of up to 160 characters to other users in a cellular network almost anywhere in the world within a short time period [18] [19]. Two types of message delivery mechanisms are used in cellular network; cell broadcast (SMS-CB) and point to point messaging (SMS-PP) [20]. In cell broadcast, a message is sent to all the active users or mobile stations (MSs) present in a specific cell. This service is only one way and there is no need for any receipt message confirmation. This service is typically used for sending real time updates such as traffic conditions, weather updates, stock market updates, and sports results and so on [7]. However, In point to point messaging, messages can be sent from one Mobile station to another mobile station. P2p message delivery mechanism provides guaranteed delivery of messages between mobiles via acknowledgments.

In [10] , the authors give a detailed overview of Skype peer to peer internet telephony protocol architecture and its functionality. The basic skype architecture is composed of three main components includes centralized login server, ordinary node or skype client, and super nodes. Skype uses centralized server for user registration and authentication while ordinary nodes are Skype client applications which enable phone calls and sending messages. Super node-based peer-to-peer networks organize participants into two layers: super nodes, and ordinary nodes. Typically, super nodes in skype protocol maintain an overlay network among themselves, while ordinary nodes pick one (or a small number of) super nodes to associate with; super nodes also function as ordinary nodes and are elected from amongst them based on some criteria. Ordinary nodes issue

---

[2] http://www.gsmworld.com/Technology/sms/intro.html.

queries through the super node(s) they are associated with. We are basically inspired from the Skype protocol but our network architecture and message distribution mechanism is different from the Skype peer to peer protocol. Skype protocol always requires network connection for communication but our proposed system don't require any network connectivity. We are using short message service as data carrier. Skype uses super nodes to manage users and form overlay network for communication while we are using hierarchy based routing for delivering information contents. In hierarchical routing, first we group similar users into clusters based on interest with their geographical vicinity, selects a cluster head for each new cluster, and then deliver required data to users based on their cluster membership through their cluster head nodes via SMS.

The study in [11] analyzes a prototype system which is capable of sharing real time stock updates using GSM short messaging service. The prototype system is basically based on client-server architecture model; consist of a web based stock price server and a gateway PC system. Centralized server connected with both the internet and the GSM cellular network. All the mobile terminals are subscribed with system through centralize server. The prototype system allows subscribers to receive update messages from the central server in different interval of time or request/reply based. In addition, the prototype system also provide an enhance delivery service called dial up interactive voice response system (IVR). This delivery service enables mobile subscribers to receive immediate update using voice prompt instead of sending request via SMS. In prototype system, almost all the functionalities related to message delivery, subscribers information management, and IVR services are performed by single centralized server. Therefore it may cause the typical problem of bottleneck when there is large number of subscribers and their instantaneous requests for content information. Therefore the prototype system may have a single point of failure and also have a possibility of delay in content delivery. With these issues, the proposed system does not provide any location base service.

In study [12] discussed a suitability of SMS for emergency warning systems. The study described that short messaging service can be used to transmit messages to a large number of mobile users in order to warn them if there is any emergency. The delivery of these messages is based on the best effort basis and does not provide any guarantee for message delivery. The study also explores the message lost probability and transmission time over the GSM network. Then the paper present a solution which is based on cell broadcast in GSM and also showed numerical results that SMS can be useable for emergency cases. The proposed system is a good effort for the smaller number of mobile users but the system may fail to achieve the primary goal of delivering messages in case of any emergency if there are large numbers of mobile users and they need to receive messages without any delay. The system is design on the basis of centralized component architecture, so the whole system functionality depends on a single component.

The paper [13] discussed a network layer protocol for wireless sensors networks (WSNs) called SGcast. The protocol allows the sensor nodes to send message to all of the nodes in a specific geographical area without having any prior knowledge of recipient nodes. The protocol minimizes the radio transmissions and better handles traffic controls in order to save energy in sensor nodes. The author in this paper introduced a novel approach of using negative hop to find destination region instead of defining target zone. The protocol also used a dynamic heuristic instead of using static heuristic for deciding whether the Geo-cast packet is forwarded or not. The protocol also introduced its own back off mechanism and provide guaranteed hop-by–hop acknowledgement.

Geographical routing is the best routing scheme for the wireless sensor networks (WSNs) because all WSNs have several nodes; all these nodes have limited power [14]. The study in [14] focused on how to design a routing protocol to save energy consumption in WSNs. Author proposed a load balance aware and energy efficient geographical routing protocol named BEGR. The BEGR protocol considers both the location and energy of the nodes in WSNs. The protocol basically assigns a cost value to each node in WSNs; the cost value is depending on the location and energy of a node. The protocol then selects the next node on the basis of minimal cost. In this way, the BEGR protocol always selects the node which has more energy as compared to other nodes. So the energy consumption is equally distributed in different nodes over the whole network.

In reference [15] author present a routing protocol for urban mobile networks based on geographical location. The protocol uses current location information and route of vehicles in mobile networks. The protocol is basically focused on minimizing the number of messages required to send messages to the destination. It also considers the time interval required to send messages between different nodes in mobile networks. The proposed protocol is based on the “store carry and forward paradigm”. The routing algorithm uses locally available maps and trajectory information of different vehicles that they near to the destination node. The proposed protocol is suitable for the limited number of vehicles, and minimum wireless range. It may not give the same performance if there is greater number of vehicles with a high speed and vast wireless range is available in routing environment.

Geo casting is the way of delivering messages from sender to the destination present in a specific geographical region [16] [17][18]. For geo casting, there are many Geo-cast protocols are proposed especially for wireless sensors networks and ad hock mobile networks. In the paper [19] author discussed different geo cast routing protocols. These geo cast routing protocols can be differentiate on the basis of flooding, direct flooding or without flooding approaches used for routing in networks. The geo cast protocols also differ based on their suitability, either they suitable for ad hock networks or they more suitable for infrastructure networks. On the basis of above mentioned parameters, author categorized the geo cast routing protocols into three basic groups; flooding, direct flooding, and no flooding. Flooding and directed flooding approaches are used for ad hock networks while the no flooding approach is suitable for both ad hock and infrastructure networks [19].

In reference [16] author presented an efficient geo casting protocol for large scale wireless sensor networks which support multiple regions. The protocol known as recursive multi-region geo-casting (RMG) protocol, it transmits a message from source node to multiple geo cast regions in large scale wireless sensor networks. The basic idea behind this algorithm is considering a multiple geo cast regions as a destination point and forward the message to all these regions as a group. The recursive multi-region geo casting protocol is performed by each node for every sending packet in large wireless network; therefore it takes large computation time on every node. Although the proposed protocol calculate division point and also required to make sub groups from a group of destinations, it may increase the complexity and delay time to transmit a message. It may also affect the quick delivery of messages to its destination.

The study [20] described an application scheme of published/subscribe system for ad hock mobile networks. The application scheme supports data distribution in clustering mobile ad hock networks (MANET). Author introduced a new scheme for publish/subscribe systems for clustering in MANET. Every cluster in MANET has information about the topics of publish/subscribe system for matching. Every cluster head in MANET is responsible of establish communication between publisher and subscriber nodes [8]. The system architecture is consist of three main components includes publisher, subscriber, information repository and four categories of objects; notifications, publications, subscriptions, and acknowledgements. The publisher in this model provides information for publication; the subscriber shows their interest in subscription, while the information repository provides the common interests by matching between publisher and subscriber. While in mobile ad hock network, there are many clusters and each cluster has one head node. These cluster head nodes are responsible for divide the MANET into different cluster [21]. The cluster head node also performs the functionality of IRN therefore the IRN node is basically a cluster head node in this architecture. Although the proposed architecture is simple and real time scheme for delivering information but it basically depend on the functionality of cluster head nodes. Therefore, the proposed system may increase the overload on the IRNs and also require a high performance of IRNs.

The study [22] discussed a tree based geo cast routing algorithm for connecting heterogeneous networks. In the proposed algorithm, a source which is a node in mobile network sends data packet to the group of nodes in other network. The algorithm works on the basis of constructing tree structure according to the route. Tree structure is formed while considering different parameters of the network such as power and direct connectivity of neighbor nodes. Before constructing a tree for route, algorithm sends hello messages to updating the list of neighbors. Once a tree structure is formed then the algorithm sets the link life time and also periodically checks link life time in order to maintain the link state along the route. Described algorithm is an efficient and robust method for heterogeneous networks but route discovery and constructing tree structure for route may cause delay in message delivery. Every time algorithm needs to find route and construct a route tree due to allowing node mobility. If a link fails or a new path to appear in the network then it requires rebuilding the path and also needs to maintain the life time of a link.

2.2. Summary

Although many researchers have already discussed geographically distributed algorithms more specifically for wireless sensors networks (WSNs) and Mobile ad hock networks (MANETs) by using different communication technologies and methods. In these contributions, to find stable path and deciding best routing strategy researchers mostly consider the energy, link up time, current geographical location, mobility, and current status of nodes in both WSNs or MANETs [13] [23] [24]. A very relevant work have been done in reference [25] [11] [12] for cellular networks utilizing short message service as a message delivery mechanism. These proposed systems are based on the client-server or centralized server model; in which centralized server is responsible for doing almost every task related to timely delivery of messages to various subscribers. These systems efficiently and timely deliver latest information contents to subscribers in normal conditions through SMS. However, traffic blocking, delay in message delivery, server overloaded issues, congestion problems in network, lack of scalability, and single point of failure are some of the basic problems related to these architectures. It may happen when mobile subscribers increase rapidly in network and large number of simultaneous requests from subscribers to the centralized server (content provider). In order to overcome these typical issues of centralized server architectures, the system should be design in a purely distributed fashion on the basis of subscriber’s current geographical location. The functionality of centralized server may distribute across many nodes in cellular network. Some of the subscribers' mobile devices in wireless cellular network should take part in message delivery process in order to increase the robustness of network. These mobile devices not only limited to receive messages from server but also save, make a copy, and further send copy of received message to other mobile subscribers based on their group membership.

Chapter 3
Design and Implementation

The previous chapter described about the background and literature survey required for implementing the proposed system. This chapter gives methodology an outline of the proposed system architecture and its implementation. The first section explains the overall system architecture and its components. The second section described the proposed algorithms related to user clustering and message distribution applied in the experiments in next chapter. It also explains how the system is implemented and how the final product works. We used hierarchical clustering approach in our research for making clusters in different geographical regions. Finally, the chapter will end with summary.

3.1. System Architecture Overview

A client application for Android mobile platform was developed. To subscribe service users' needs to install application on their mobile phones. The application fetches user's mobile current position information from the GSM cellular network using Android location API. At the time of initial registration subscribers set their preferences including name, interest, their device capability (maximum number of messages sends by a user) and some other information as required by the service. The client application then combines user preferences with location information of mobile devices, and application sends this information to a centralized web server through short message service for service subscription.

At the server side, a GSM modem is attached with centralized server which receives SMS from subscribers and forward to the application. At the second step, application server creates user's profile and group subscribers to make clusters based on their interest with geographical vicinity. After clustering server elects a head node for each cluster and sends messages to these cluster heads via SMS. The cluster heads in different geographical regions receive SMS and further forward to their cluster members. Eventually, each subscriber receives required information contents from the content provider (centralized web server) in a full geographically distributed computing environment.

Skype peer to peer telephony protocol [25] and multi-level hierarchical clustering [26] [27] of similar objects have been our major inspiration for design of the proposed system. So the proposed system is divided into three components which include:

- Centralized Application Server.
- Network Relays or Intermediate Processing Nodes.
- Mobile Application.

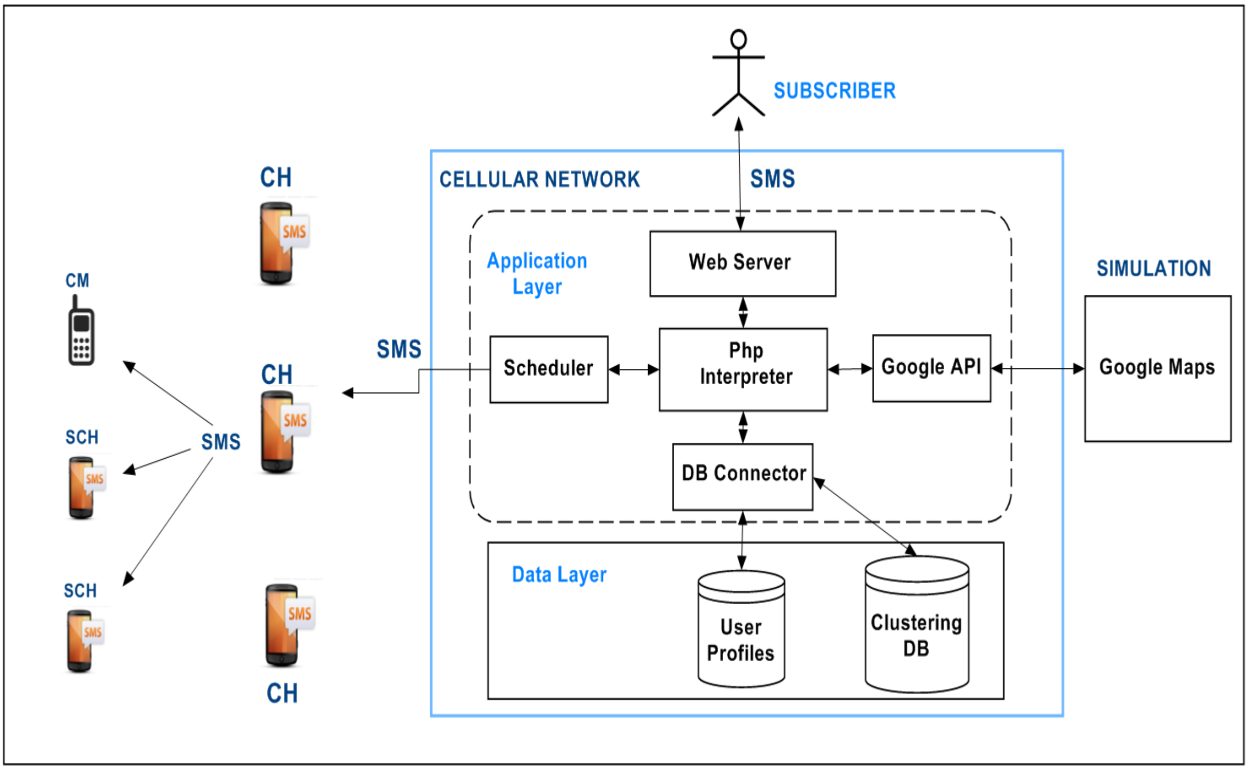


Figure 3.1 System Application diagram

Figure 3.1 shows overall system application diagram in which prototype operates. Each component of the system performs its own individual functions and is dependent on each other. Next section explains all of these three components in detail. Fig 3.2 represents complete system architecture.

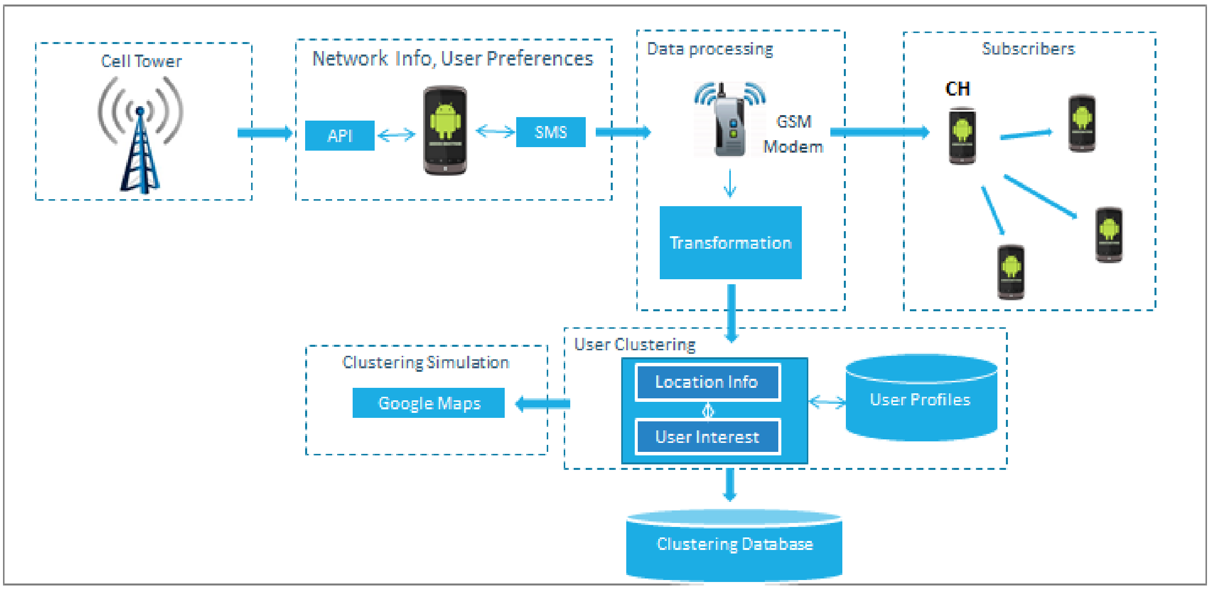


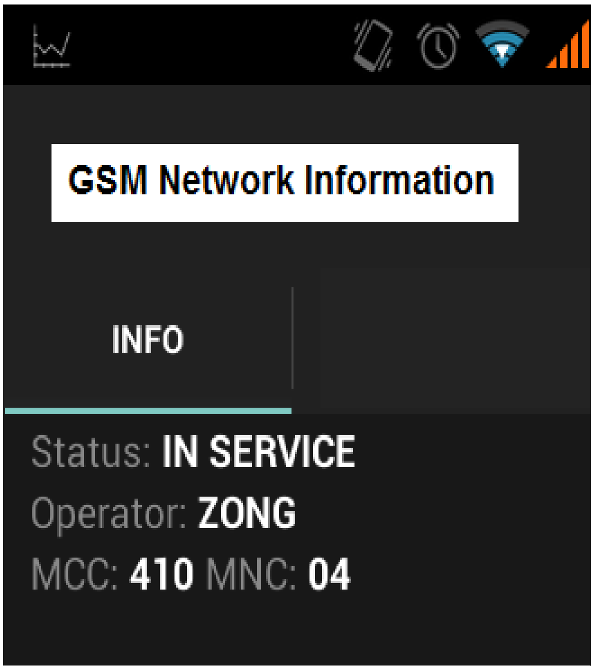
Figure 3.2 System Architecture

3.1.1. Centralized Application Server

The centralized web based application server acts as central repository and the source (Contents Provider) to various registered subscribers. There is obviously no restriction on where the centralized application server is located anywhere geographically. The application server receives registration requests from subscribers through short message service (SMS). The server creates a profile for each new subscriber and performs clustering to create a multi-level hierarchy in different geographical regions. Subscribers are grouped into clusters according to their interest with current geographical location. After clustering server also elects head node for each cluster and maintains a database for subscribers clustering. Finally, the centralized application server sends latest information contents to group of subscribers through cluster heads via SMS.

3.1.2. Data Collection Module

The mobile application is the client application that would install and run in the subscribers Android mobile phones. The initial functionality of the mobile application is to determine the current geographical location information of the subscriber and to display the received information contents on subscriber mobile device. The sample subscriber current geographical location information is shown in figure below.



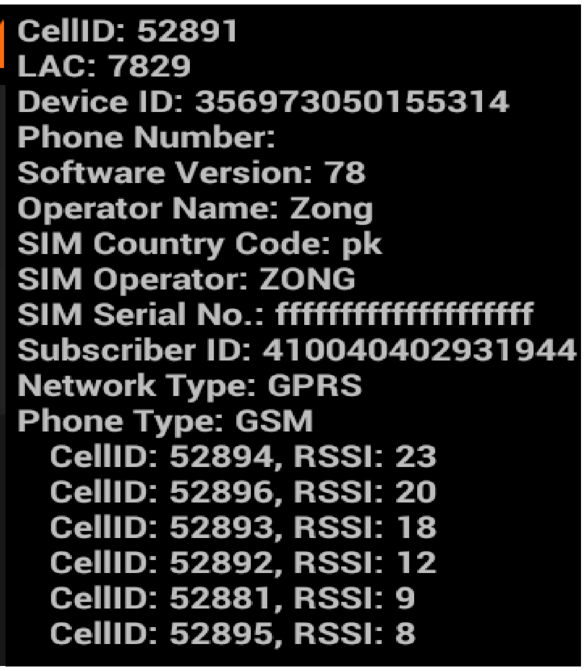


Figure 3.3 Network information

Figure 3.4 Neighboring Cell IDs

The data collection module or mobile application collects data from mobile subscribers and GSM network and sends to the centralized application server for subscription. Every user who wants to subscribe services first needs to install application on their mobile device. It has location information fetching module through which it determines subscribers' current position information using cell-id based method. Fig 3.3 and 3.4 both shows GSM network information which includes current tower cell-id, mobile operator, location area code (LAC), and some other information related radio signals. In order to collect data from subscribers' preferences, application provides graphical user interface (GUI) through this interface subscriber set their preferences. The application then combine both location and subscriber preferences information in a message and sends it to the server for registration via SMS over cellular network.

The mobile application receives messages directly from the server or through network relays (cluster/sub-cluster heads) via SMS. After receiving messages it can store, make a copy of received message, and further forward to other cluster members or sub-cluster heads. Finally, it displays actual information contents of subscriber interest on their mobile devices.

3.1.3. Network relays or Intermediate Processing Nodes

The third and most important module of the system is intermediate processing nodes or network relays. The system used some of the subscribers' mobile devices as network relays for message distribution. Whenever a new cluster or sub-cluster is formed one of the mobile device is selected as head node for that cluster. Through these nodes server implements both user clustering algorithm and message distribution mechanism in network. These nodes are called network relays or cluster and sub-cluster heads.

There are also some other nodes in network which are not working as network relays these node are referred as ordinary nodes or cluster members (CM). Overall message distribution process will have three types of mobile nodes, all maintaining at different levels with different responsibilities, which includes:

- **Cluster Head.** These are ordinary mobile devices in network with some additional functionality. Cluster heads are the intermediate processing nodes which are directly connected with centralized web server. They are located at the top level hierarchy so receive both control and data messages directly from the server and further forward to its remaining cluster member nodes via SMS.
- **Sub-cluster Head.** These are also intermediate processing nodes under the cluster head nodes. These nodes are located under the cluster head and create second level hierarchy in network. Sub-cluster heads are functioning under the top level cluster heads therefore they receive data messages from their upper level cluster head and further forward to its sub-cluster member nodes.
- **Ordinary Mobile Node or Cluster Members.** These are the ordinary mobile devices in the network**.** These nodes are not functioning as an intermediate processing node therefore these nodes have no additional functionalities of cluster implementation and message distribution. Ordinary mobile devices in network are must be a member of any cluster or sub-cluster and limited to only receive information messages through its upper level head node.

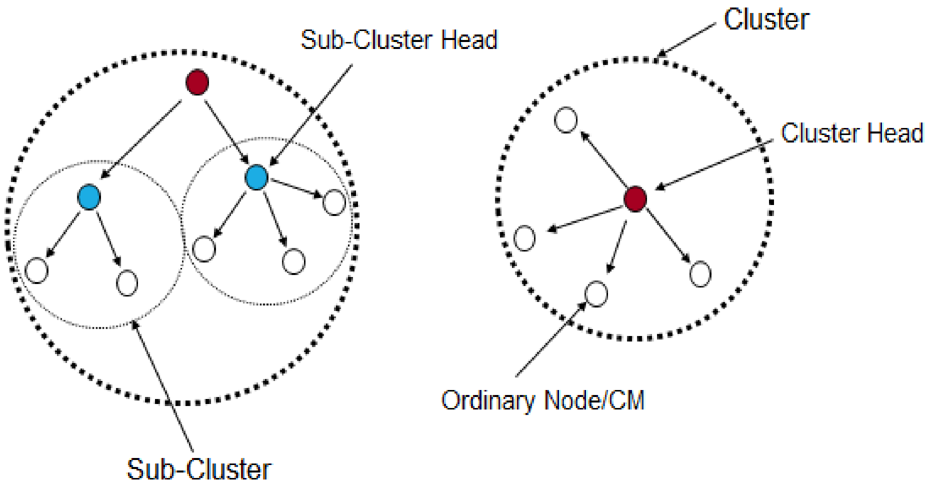


Figure 3.5 Cluster Hierarchy

Fig 3.5 shows overall network hierarchy, clusters, sub-clusters, and ordinary nodes/cluster members in cluster hierarchy.

3.2. System Implementation

Implementation of the system can be divided into three further phases. First phase consist of development of a location-based cluster hierarchy, where every new subscriber either added into the existing cluster or create a new cluster by adding new subscriber without considering any improvement in existing hierarchy.

The second phase introduced cluster organization algorithms for handling size of mobile subscribers in different geographical regions. These algorithms dynamically split and merge existing clusters on the basis of current size of subscribers. So these algorithms ensure the placement of similar and geographically closer mobile subscribers in same cluster or sub-cluster in order to improve cluster hierarchy.

The third phase of implementation introduced an algorithm for cluster head selection as well as for sub-cluster head which results in selection of network relay nodes for message distribution. This phase incorporates the efficient node search for cluster head selection at the time of each new cluster or sub-cluster formation. It also save information regarding other cluster members and maintains the changes in cluster hierarchy at the removal and arrival of mobile subscribers in network.

The fourth and last phase of the implementation process introduced a geographically message distribution mechanism with the help of intermediate processing nodes also known as network relays. This phase can be further divided into three sub-phases according to responsibilities. These three responsibilities are initial service subscription or subscriber registration phase, cluster-based message distribution, and mobility management in network. This phase is overall responsible for message delivery from centralized server to various mobile subscribers according to their cluster membership. The phase uses all the algorithms as we introduced and implemented in above three phases utilized for the implementation of described responsibilities.

All the proposed algorithms as mentioned in above first three phases of implementation are described in next section while the fourth and last phase of implementation process with its responsibilities are described in the last section of this chapter.

3.2.1. Hierarchical or Cluster-Based Routing Algorithm

The hierarchical clustering algorithm makes it possible to divide the mobile subscribers in cellular network into some clusters and sub-clusters according to their current geographical location. Each subscriber must be a member of at least one cluster and each cluster and sub-cluster will have a cluster head that would be responsible for message distribution within each cluster. So we have implemented new algorithms for both hierarchical user clustering and cluster organization as described below.

3.2.2. Hierarchical User Clustering

We proposed a new clustering algorithm for grouping mobile subscribers into clusters to form a multilevel network hierarchy. At any level the proposed algorithm can make clusters and sub-clusters of different densities and sizes based on some similarity criterion. The clustering parameters and number of clusters and sub-clusters in different geographical regions are decided automatically for given datasets. The clustering algorithm starts with calculating distance between two geographical points (basically represents current location of mobile subscribers) and a geographical point to existing clusters. Geographically nearest mobile subscribers are then grouped to make clusters. The subscribers to be clustered are basically a set of geographical points with latitude and longitude values on which an appropriate distance matrix ***d*** is defined. The input to the proposed algorithm is a threshold value for maximum size of a cluster with proximity matrix ***d*** and the algorithm output is the multilevel hierarchy of clusters in different geographical regions.

Dynamic hierarchical clustering algorithm (DHCA) grouped similar users into clusters according to their interest and geographical vicinity. At start DHCA group users together into clusters based on the geographical distance from each other. For example group all users inside 30 kilometer geographical radius to form a cluster. After that cluster in different geographical regions are created based on the distance between a new user and existing clusters (centroids). Cluster centroids are normally determined algorithmically through repetition of the existing users locations (latitude & longitude values). The geographical distance between users can be determined in several ways[3]. In order to compute distance between two geographical points on earth we used Haversine formula. Haversine formula[4] is practically accurate and commonly used. It presumes that earth is sphere-shaped (in reality earth is marginally ellipsoid) and uses spherical geometry to compute the distance between two geographical points on the sphere. This affects accuracy to be +-2 kilometer when computing distances of about 20.000 km. It used 6371.0 kilometer as average radius of earth. The Haversine formula is mathematically equivalent to the Law of Cosines, but is frequently preferred since it is less sensitive to round-off error that can occur when measuring distances between geographical points that are located very close together. The Haversine formula is given in the following equation.

$$\text{haversin}\left(\frac{d}{r}\right) = \text{haversin}(\phi_2 - \phi_1) + \cos(\phi_1)\cos(\phi_2)\,\text{haversin}(\lambda_2 - \lambda_1)$$

In the above equation $\boldsymbol{\varphi_1}$ and $\boldsymbol{\varphi_2}$ are latitudes of geographical Point 1 and Point 2, $\boldsymbol{\lambda}$ represents longitude value, $\boldsymbol{\Delta\varphi = \varphi_1 - \varphi_2}$ is latitude separation, $\boldsymbol{\Delta\lambda}$ is longitude separation, $\boldsymbol{r}$ is the average radius of earth, and $\boldsymbol{d}$ is the required distance between two geographical points on earth.

Similarly in order to find distance between two clusters in network hierarchy we used shortest distance formula as given in the following equation.

$$D_{centroids}(C_i, C_j) = d(r_i, r_j)$$

[3] http://www.sunearthtools.com/tools/distance.php.
[4] http://www.longitudestore.com/haversine-formula.html.

In above equation $\mathbf{C_i}$ and $\mathbf{C_j}$ are two clusters and $\mathbf{r}_i$ and $\mathbf{r}_j$ represents the center of cluster $\mathbf{C_i}$ and $\mathbf{C_j}$ respectively. While ***D*** is the required distance between the center $\mathbf{r}_i$ of $\mathbf{C_i}$ and the center $\mathbf{r}_j$ of $\mathbf{C_j}$.

Finally, we used minimum distance formula[5] in order to compute geographical distance between a point $\mathbf{P_1}$ and an existing cluster $\mathbf{C_i}$. The minimum distance formula is given in equation below.

$$D_{sl}(C_i, C_j) = \min_{x,y}\{d(x,y) | x \in C_i, y \in C_j\}$$

The steps for hierarchical clustering algorithm are as follows:

Clustering Algorithm
(Repeat Process for each new user) 1. Select new user with their preferences U_i = {***lat long, i, CID,*** d_c}. 2. Create user profile on the basis of their preferences. 3. Select all the existing clusters **C= {*C₁, C₂, C₃……, Cₙ*}** with their members. 4. Select predefined radius ***r***. 5. Compute distance *d* from the location of new user to all the existing clusters ***C*** lies within the within the predefined threshold value of ***r***. ***6. IF*** 7. There is only one cluster is found within the given radius ***r*** then adds the new user into the current cluster. **8. *THEN*** 9. If more than one cluster is found within given radius ***r*** now select the cluster with shortest distance ***d*** and add new user U_i to that cluster. **10. *ELSE*** 11. If there is no any cluster exists within the given radius ***r*** at that time forms a new cluster by adding current user. 12. Assign a unique identifier to each new cluster. 13. Update the clustering database. 14. Repeat all steps for each new user at registration. **(Compute distance *d* by using Haversine distance formula)**

Algorithm 3.1 Dynamic Hierarchical Clustering

The centralized server based dynamic hierarchical clustering process consists of three distinct levels; interface level, algorithmic level, and index level. As shown in the following figure.

[5] http://www.codeguru.com/cpp/cpp/algorithms/article.php/c5115/Geographic-Distance-and-Azimuth-Calculations.htm

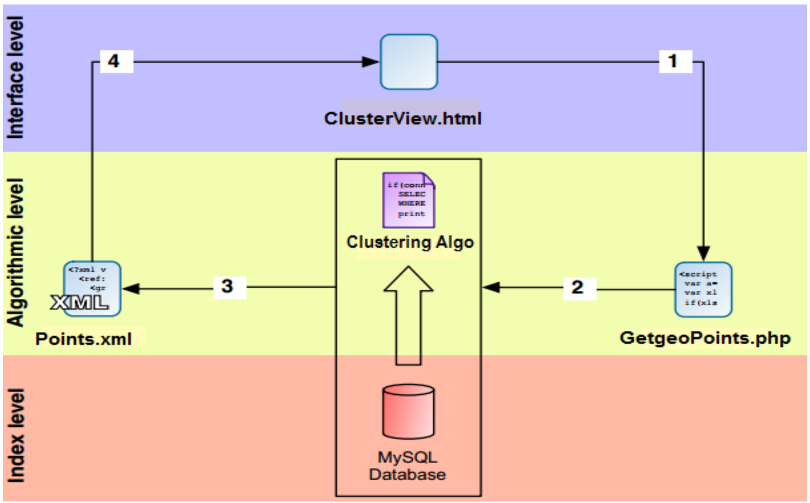


Figure 3.2Hierarchical Clustering Architecture

3.2.3. Dynamic Cluster Splitting

The dynamic cluster splitting algorithm is dependent on the parameter ***s*** which is the threshold value for maximum subscribers in a cluster at any level. If the number of subscribers in a cluster is increased to the threshold value ***s*** then this cluster is divided into two smaller sub-clusters. The steps for dynamic cluster splitting algorithm are as follows:

Cluster Splitting (Max User, Radius)
Input: **C1, C2, C3** ……………**Cn**. // Set of Clusters **S** // Maximum Users in a cluster **Output:** Multiple Sub-clusters C_{i1} , C_{i2} **Process:** 1. Select a cluster whose size is exceeded with threshold value. 2. Divide cluster into two smaller balance cluster C_{i1}, C_{i2}. 3. Label each sub-cluster C_{i1}, C_{i2}. 4. Update level of the clusters. **5.** Repeat steps **1** to **4** for each Cluster.

Algorithm 3.2 Dynamic Cluster Splitting

Note that only one cluster can split but can never merge at a time in network hierarchy. Whenever a new subscriber is added to an existing cluster in network hierarchy, the centralized application server will always check predefined threshold ***(s)*** value. If the number of subscribers is increased from the threshold value then the cluster splitting algorithm is executed by centralized server in order to make sub-clusters.

3.2.4. Dynamic Cluster Merging

The dynamic cluster merging algorithm depends on the parameter ***t*** which is the threshold value for minimum subscribers in two sub-clusters under the same cluster head at any level in network hierarchy. If the number of subscribers in a cluster is decreased to the threshold value ***t*** then these two smaller sub-clusters are merged to form a single larger cluster. Note that only two sub-clusters under the same cluster head can merge at a time. Whenever a mobile subscriber leaves the cluster for any reason the centralized application server will always check the threshold ***(t)*** value. If the number of subscribers in both sub-clusters is less than the threshold ***(t)*** value then the cluster merging algorithm is executed by centralized server in order to make a large single cluster by merging small clusters.

The steps for dynamic cluster merging algorithm are as follows:

Cluster Merging

Input:
C_i , C_j ,……..C_n // *Set of sub-clusters*
threshold value **t** // *Least number of users in two sub-clusters*
S // *Current user size*
Output:
$\mathbf{C_n}$ // *Large Single Cluster*
Process:

1. Among the current sub-clusters, determine the two sub-clusters c_i and c_j under a same cluster head (that are most nearest).
2. Find total number of users in both sub-clusters & compare with predefined threshold value.
3. IF **S** < **t**, merge C_i and C_j into a single cluster **C** = $C_i \cup C_j$.
4. Label & update the level of the new cluster $\mathbf{C_n}$.
5. Elects head node (**CH**) for new cluster.
6. Repeat the process for each pair of sub-clusters under a main cluster whenever a user leaves or unsubscribe the service.

Algorithm 3.3 Dynamic Cluster Merging

3.2.5. Cluster Head Selection

In cluster based routing algorithm, it is essential for each cluster and sub-cluster to have a head node in order to implement clustering, message routing decisions, maintain and control the network hierarchy in a fully distributed fashion. Cluster head is also liable to make sure that each and every member of the cluster timely receives messages. In the proposed algorithm, cluster head (CH) has to do extra job for sending information contents via SMS to other cluster members or sub-cluster heads, therefore maximum number of SMS are needed by cluster head to deliver information. So in order to timely and successful message delivery it is most appropriate to select a node as cluster head which has maximum number of SMS. Consequently the head selection algorithm used message sending capacity (device capability) of a user as a performance parameter for the selection of cluster head. The message sending capacity is defined by each subscriber at the time of initial service subscription.

We introduced a new cluster head selection algorithm for wireless cellular network. The cluster head selection algorithm elects any one subscriber's mobile device as a head node for each cluster and sub-cluster based on the criteria as described above.

The steps of cluster head selection algorithm are as follows:

Cluster Head Selection Algorithm
Input: **C1, C2, C3 ………Cn.** // Set of Clusters with their members d_C // Message sending capacity of each device **Output:** **CH** // Cluster head with highest d_C. **Process:** 1. Select cluster or sub-cluster with their members. 2. Determine maximum device capability d_C. 3. Select device which has greater value of d_C. 4. Update the clustering database on server. 5. Notify newly elected cluster head (**CH**). 6. Repeat process for each cluster and sub-cluster in the network.

Algorithm 3.4 Cluster Head Selection

Clusters and sub-clusters are formed in a way that every cluster has a minimum of one cluster head node which is responsible for getting information contents from centralized application server and further forwarding it to its remaining cluster member nodes or next level sub-cluster heads. This node is cluster head in a sense that it is directly communicating with centralized server and receiving required information contents. It is also responsible for delivering these information contents to other cluster members. Cluster head node does not suffer with extra burden to send messages to its cluster members because subscribers are well distributed according to the size in different geographical regions. If a cluster has too many subscribers, it may full by reaching the maximum number of subscribers (exceeding the threshold value). In this scenario existing cluster is further divided into two smaller sub-clusters and server elects head node for each sub-cluster. So the cluster head now needs to forward messages to sub-cluster heads only. And each sub-cluster head receives messages from its upper level head node and further forward to its cluster members. In this way, a cluster head may have one or more sub-cluster heads that are also acting as network relays to deliver information contents to all subscribers of that cluster.

Head node selection algorithm is executed by centralized application server in the following scenario:-

1. For the first time when the cluster-based hierarchy is implemented, the selection of cluster head is done by centralized application server.
2. Whenever a new cluster or sub-cluster is formed then cluster head selection algorithm is done by the server.
3. Whenever any of the existing cluster or sub-cluster head is gone or not working properly with any cause then CH selection algorithm is run by server.
4. Whenever a cluster is merged or split then cluster head selection algorithm is executed by server.
5. Whenever a device capability (d_c) parameter of existing cluster or sub-cluster head is equal to predefined threshold value then cluster head selection algorithm is executed by server.

The above algorithm elects any subscriber as a head node for cluster or sub-cluster. So the head nodes are normal user mobile devices in cellular network with some additional responsibilities as assigned by the system to dynamically implement clustering and message distribution service. The additional responsibilities of head nodes or network relays are given below:

1. Directly communicate with centralized server for both clustering and message distribution.
2. Implement user clustering wirelessly linked with centralized application server.

3. Responsible for message delivery to targeted mobile subscribers within a cluster.
4. Maintains network topology (Clustering Info).
5. Acts as network relays receives information contents from server and further forwards received message to its cluster members.
6. They also forward messages to sub-cluster heads if any exist in network.
7. Aware of its all cluster members or sub-cluster heads.
8. Capable to receive, send, copy, save, and display message contents on mobile devices.

3.2.6. Difference between cluster & sub-cluster head

Sub-cluster head (SCH) nodes also act as a network relay just like a cluster (CH) does but the only difference is that they only receive control messages directly from centralized web server while data messages are received from the upper level cluster head. In addition, Sub-cluster head (SCH) only store their own cluster member's information and don't need to maintain a list of any cluster or other sub-cluster head nodes.

3.3. Responsibilities of the System

There are three key responsibilities of the proposed system as discussed in above section. These responsibilities are initial service subscription, message distribution mechanism, and mobility management or service assurance.

3.3.1. Service Subscription or Subscriber Registration

Initial subscriber registration or service subscription algorithm start by installing the client application on subscriber's mobile device. Once application is installed on mobile device by interested subscriber then it acquires current geographical location of subscriber from the existing cellular network. At the time of initial installation subscriber prescribed their preferences such as their topic of interest, device capability (message sending capacity of mobile subscribers), and may be some other related information. Once subscriber defined their preferences and current geographical location of mobile device has been determined, the application then sends a request message to centralized server for service subscription by using short message service (SMS) over cellular network. We have attached an Android mobile phone with centralized application server which provides the GSM gateway functionality. The centralized server receives service subscription request message from every new mobile subscriber through attached GSM modem. After receiving service subscription messages, centralized application server first extracts subscriber preferences with their geographic location information and creates a profile for each new subscriber in database. Once user profile is created in database the next step is to find a cluster for the new subscriber according to their interest with current geographical vicinity and it is done by hierarchical clustering algorithm as explained in section 3.2.2. If existing Cluster is not found within a defined geographical radius then server formed a new cluster by adding this subscriber as cluster member and updated clustering database. So initially new added subscriber is by default selected as a cluster head and starts direct communication with centralized server via SMS.

Subscriber registration or initial service subscription algorithm explained through a flow chart diagram as shown in figure 3.6.

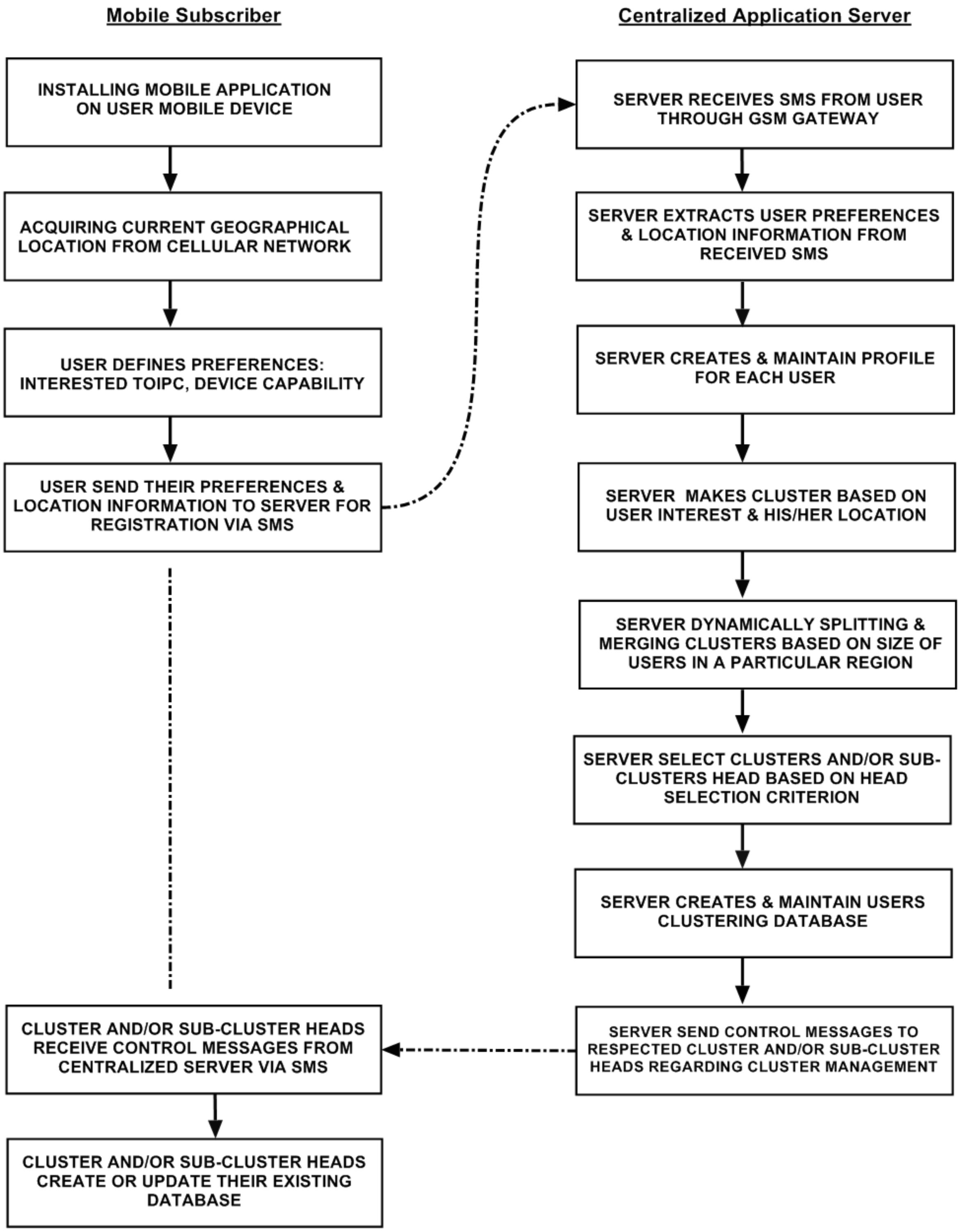


Figure 3.6 Flow Diagram for Initial service subscription

3.3.2. Message Distribution Mechanism

Message distribution phase used cluster-based message routing algorithm to deliver information contents to a group of mobile subscribers based on their membership to cluster or sub-cluster in network hierarchy. The main objective of the hierarchical cluster-based message routing algorithm is to distribute load among nodes and minimize message communication cost. In cluster-based routing, cellular network is divided into several clusters and sub-clusters according to the size of users in different geographical regions. Each cluster and sub-cluster is managed by the centralized server with the help of network relay nodes. There are three types of nodes in network hierarchy cluster head (CH), sub-cluster head (SCH), and cluster members (CMs). Every cluster and sub-cluster head have information about its member nodes. Head nodes in cluster hierarchy acts as network relay that receives short messages from server and further forward these messages to other cluster members or sub-cluster head nodes. Cluster members only receive messages from head nodes and display to subscribers on their mobile devices. In our proposed architecture only one head node for each cluster and sub-cluster has been supported and all cluster heads can directly communicate with centralized server. While sub-cluster heads only receives messages from upper level head node and transmits these messages to its remaining cluster members through SMS. Eventually, each and every registered subscriber in network receives information contents to their interest.

In message distribution phase, Following type of message communication in the network can take place between server and various mobile nodes and in between different types of nodes:-

1. The centralized application server sends information contents to only cluster head (CH) nodes via SMS.
2. Cluster head nodes receive messages and further forward these messages to its remaining cluster members via SMS.
3. Cluster head also sends messages to sub-cluster heads if they have sub-clusters in network hierarchy.
4. Sub-cluster head receives messages from its upper level head node and send it to other member nodes in sub-cluster via SMS.

5. Centralized application server also sends control messages to both cluster and sub-cluster heads in order to managed cluster hierarchy.

Hierarchy levels and message distribution process is shown in following figure:

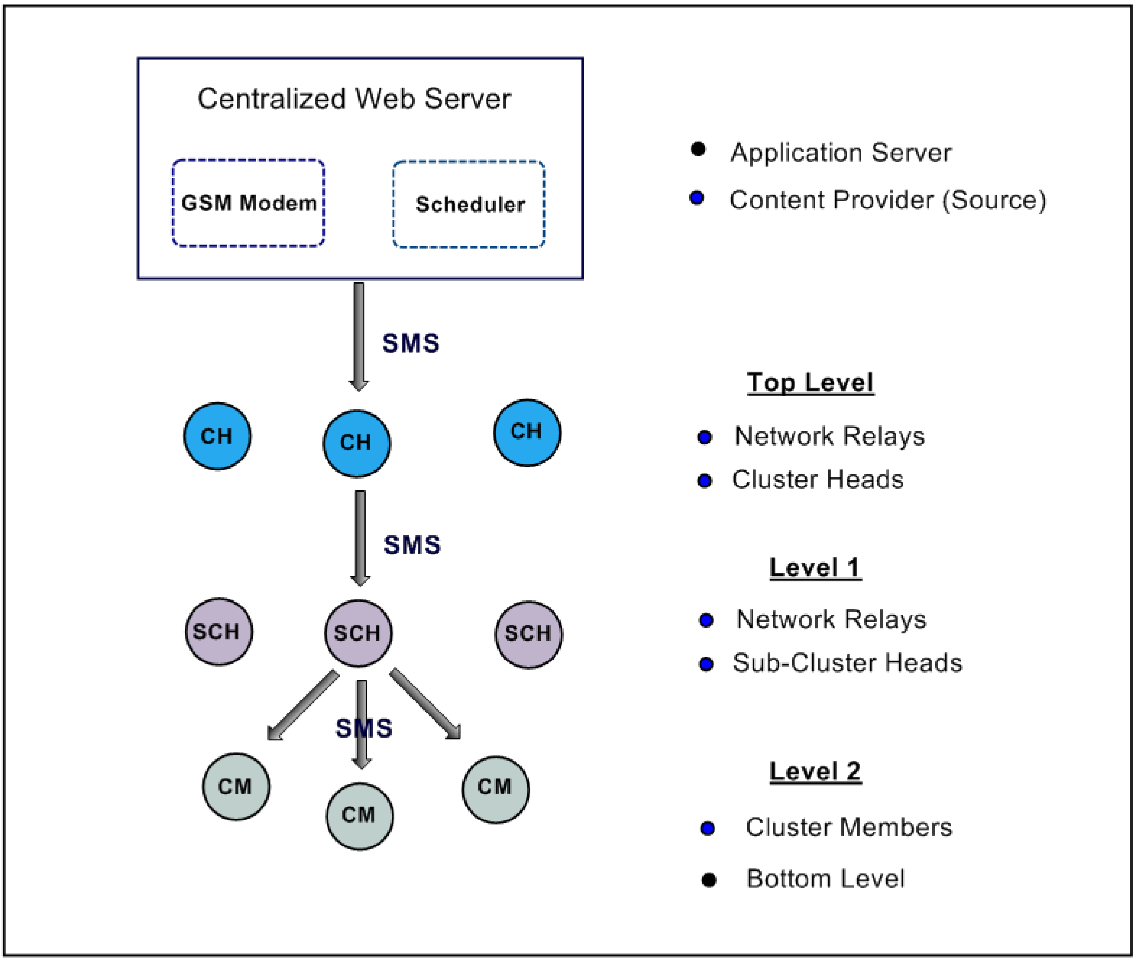


Figure 3.7 Hierarchy Levels and Message Distribution

3.3.3. Service Assurance or Mobility Management

Mobility management or service assurance phase insures the continuity and reliable message dissemination service to registered mobile subscribers if any node in cellular network drops during operation. Service assurance allows the continuity of service if a peer drops during message distribution process at any level in cluster hierarchy. We have three different types of nodes in our proposed system so the mobility management scenario is different for each type of node. All the possible scenarios are described below:

3.3.3.1. Network Relay Nodes Mobility

The cluster and sub-cluster head nodes are very important and play a crucial role for reliable message delivery to a group of mobile subscribers in cellular networks. Each cluster and sub-cluster head is used to implement the message distribution policy and maintains cluster hierarchy in network wirelessly linked with centralized application server.

- All the member nodes in a cluster or sub-cluster can monitor the respective head node through keep-alive messages to make sure that cluster or sub-cluster heads are alive and active in the network. All the cluster and sub-cluster members who do not receive the most recent message within a predefined time period can start to send error messages directly to centralized web server. If the server receives a specified number of error messages from the users that all belong to a specific geographical region than server assumes that upper level cluster or sub-cluster head is probably missing because it has the application uninstalled or switched off, or is in area without cellular coverage. After confirmation by communication with desired node through alert messages, centralized server selects a new head from remaining cluster members using head selection algorithm.
- If the cluster or sub-cluster head reached to its maximum message sending capacity then there is need to elects new cluster head and update the clustering database with new elected head node.
- If cluster or sub-cluster head moves from their initial geographical location to other location that is geographically for away from its current location and May also effect the location-based information contents. This scenario has the most destructive effect on cluster hierarchy because there is no way to further deliver messages to remaining mobile subscribers in cluster or sub-cluster. So in this case, there is need to change in cluster hierarchy in both geographical locations. Therefore centralized server in this scenario re-clusters remaining subscribers and elects new head node for the cluster and updates the clustering database on both server and head node.

3.3.3.2. Member Nodes Mobility:

The last mobility scenario is if an ordinary mobile subscriber or member node existed in a cluster or sub-cluster drops. This has no effect on the message distribution because a member node just receives messages from its upper level head node and it doesn't need to further forward messages to other cluster members. But there is need to changing in clustering database both at server and head node level. So in this case, centralized application server removes member node from its clustering database and notify to respective cluster or sub-cluster head to remove this node from its database. In this way cluster or sub-cluster heads maintains cluster hierarchy by removing dropped subscriber from their database and next time messages will not send to these dropped nodes.

All member nodes in a cluster or sub-cluster always have the same head node and they only receive messages through this node

3.4. Prototype System Design & Implementation

Figure 3.2 shows a developed prototype of the proposed system. The prototype demonstration system is designed around standard laptop system (core i5 CPU, 2.4 GHz, 4GB RAM) in windows operating system which has hardware equipped to support the variety of communication links required for the system's operation.

The solution to this prototype is to place an SMS gateway or GSM modem between the centralized server and the subscriber mobile device. The prototype of the system can then use simple GSM short message service (SMS) for sending subscribers preferences and its current location information over cellular network. For the SMS gateway we have used an Android based mobile phone as the platform to receive and send SMS messages. Somehow, we have not allocated budget for specific GSM modem. As for the modem, we have been using an Android based mobile phone (built in GSM modem). This GSM mobile phone is connected to the laptop system, which takes care of sending and receiving SMS messages over the GSM network. We have succeeded in connecting the mobile phone to our main module (centralized server) for communication. We have tested developed prototype by using Samsung Galaxy S3 and Q-Mobile A-50 as GSM modem. These devices are capable of independently send and receive messages between centralized web server and various mobile subscribers. It is a feasible solution that will allow you to start with a small number of messages for testing and validating developed prototype.

We have also developed a mobile application for subscribers' mobile devices. The mobile application was developed in Android studio 1.0.1. The mobile application is used to determine the current location of users and network information from GSM network. The location information includes current cell id (CID), location area code (LAC), mobile country code (MCC), mobile network code (MNC), and longitude and latitude coordinates. This information can be determined using the location API available in Android. A message with subscribers' preferences such as name, interest, device capacity, and their location information was sent from the user mobile device to the centralized server to a number e.g., 0312 0505944 as a text via SMS. At the server site our GSM modem which is connected to the system; received SMS, checked SMS type and its format, retrieved contents from message body, and then forward these contents to centralized server.

For implementation of centralized server which is our main application module, we have decided to code it in Php for implementing server functions. The main application module needs centralized database for the purpose of storing user profiles and clustering information therefore MYSQL was used for the database and MYSQL. We fulfilled this need by using local server on Personal computer (PC). For this purpose we used software named WAMPSERVER which include PHP, MYSQL and apache web server in one package.

3.5. Summary

All of the above described proposed algorithms have particular features to deal with specific problems of message routing and dynamic user clustering in wireless mobile networks. Because naturally, cellular network conditions (such as user size in network, message density, and users' mobility) in a cellular network are not known, it is desirable to design an algorithm that is responsive to the existing network conditions. This is why dynamic hierarchical clustering and cluster based message routing approach is chosen. The dynamic hierarchical clustering approach makes the algorithm responsive to users' mobility and network size. The relay nodes in network considerably reduce the message distribution overhead on both centralized server and other intermediate processing nodes. This results in respectable scalability to the network size. However if there is high mobility in network, user clustering must be updated more regularly to reflect the current change in network. However, when there is low mobility, it does not require changes more regularly in cluster hierarchy as the mobility does not change as much.

Chapter 4
Experimental Results and Evaluation

This chapter explains the dataset, simulation and experimental environment, and analyzes the results of executed simulations and experiments. We explained the simulation environment, and the tools used in our research. We implemented a prototype to validate our system design and evaluate the performances of proposed algorithms.

4.1. Dataset

To evaluate the performance of the proposed algorithm, we used a large collection of subscribers' datasets. The data set is collected from all over Pakistan where GSM network coverage is available. Mobile application on subscribers' devices is used as data collection module to create a richer database by collecting data from both GSM network and mobile subscribers and sending it to the centralized server. The data collected included subscribers cell phone number, mobile country code (MCC), mobile network code (MNC), cell tower IDs of different mobile operators, location area code (LAC), geographical coordinates (latitude and longitude) corresponding each cell ID, neighboring cell IDs, etc. The collected data also included information about subscriber's preferences such as their interest and device capability (maximum number of messages sends by a subscriber).

We have collected approximately 6,000 unique records in our database throughout study. All of the data is stored on centralized application server database. For our simulations we only require the data regarding subscriber's current location. This data consist of the current cell ID, LAC, neighboring cell IDs, and latitude and longitude values of each cell ID. The remaining information related to network and subscribers will use in message distribution process in next section.

The following attribute list is used by centralized application server for creating user profiles.

Attribute	Data Type
User Cell No	String
Name	String
Address	String
Cell Id	String
Location Area Code (LAC)	Timestamp
Latitude	Float (10,08)
Longitude	Float (10,08)
Users Interest	String
Device Capability	int32
MCC + MNC	int32

Table 4.1 Attribute List in User Profile

The following attribute list is used for managing clustering database by centralized application server as well as network relay nodes.

Attribute	Data Type
Cluster Id	int32
Sub-cluster Id	int32
Cluster Center	Float (10,08)
Sub-cluster Center	Float (10,08)
Users Id	int32
Cluster Head	int32
Sub-cluster Head	int32

Table 4.2 Attribute List in Clustering Database

4.2. Experimental Setup

In this section, a description about the simulation environment, tools used in simulations, and measures of performance evaluation of proposed algorithms is described.

4.2.1. Experimental Tools and Environment

The centralized application server is implemented in Php programming language on a standard computer (core i5 CPU, 2.4 GHz, 4GB RAM) in windows operating system. The server implementation also used JavaScript and google map API version 3. Mobile application is developed in java using Android studio only compatible for Android based mobile phones. We also used Android mobile phone as a GSM modem in our experiments. The database module of the implementation resides on the centralized application server. The database module stores all the data send by mobile application from subscribers mobile phones via SMS. It also stores all the information related to user clustering. We used MYSQL database due to its lightweight and high scalability feature.

4.2.2. Experimental Methodology

We have implemented a prototype system on the Android platform with different types of mobile phones and successfully send and receive messages to and from centralized server respectively via SMS. We first test the validation and performance of each proposed algorithm individually in order to evaluate the design feasibility. We start from evaluating basic clustering algorithm for different data sets in section 4.3.1. We test the hierarchical clustering and head selection algorithm in section 4.3.3. And finally we evaluate the performance of complete system in next section using developed prototype in which all proposed algorithms and system components are working collectively to achieve defined goal.

4.3. Experimental Results

We simulated and experimentally evaluated the performance of proposed algorithms to find hierarchical clusters in different geographical regions using different datasets. We also evaluated the performance of head selection and other algorithms for cluster organization. In the next section we first describe the predefined parameters with each datasets for hierarchical clustering followed by explanation of the experimental results.

4.3.1. Observing Users location with Google API

As discussed in previous section our dataset contains complete information i.e., MCC, MNC, LAC, CID, latitude and longitude values. So CID, LAC, and longitude and latitude information can be used for the users' location tracking before applying clustering algorithm. However during our experiments we observed that geographical coordinates (latitude and longitude values) against each cell ID is very important in order to track current location of users in different regions.

Fig 4.1 Display 100 users on Google map before Clustering

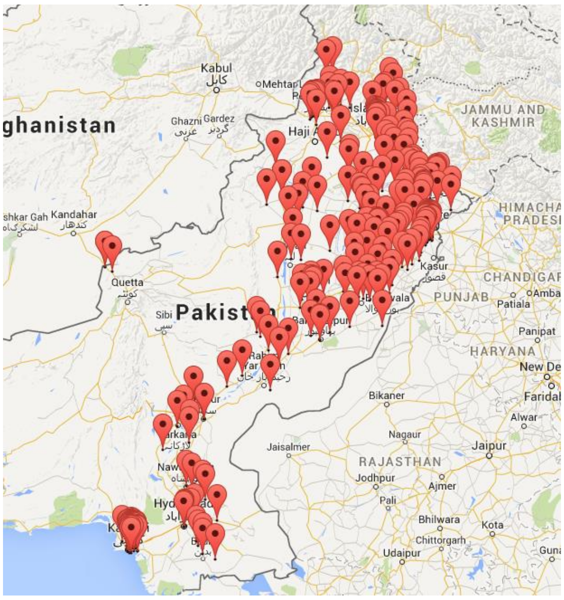


Fig 4.2 Display 300 users on Google map before Clustering

We have used a real dataset of subscribers for simulations on google map consist of 100 to 6,000 cellular subscribers. Before applying proposed clustering algorithm on different dataset first we need to analysis the current geographical location and density of mobile subscribers in particular regions of Pakistan by displaying it on Google map.

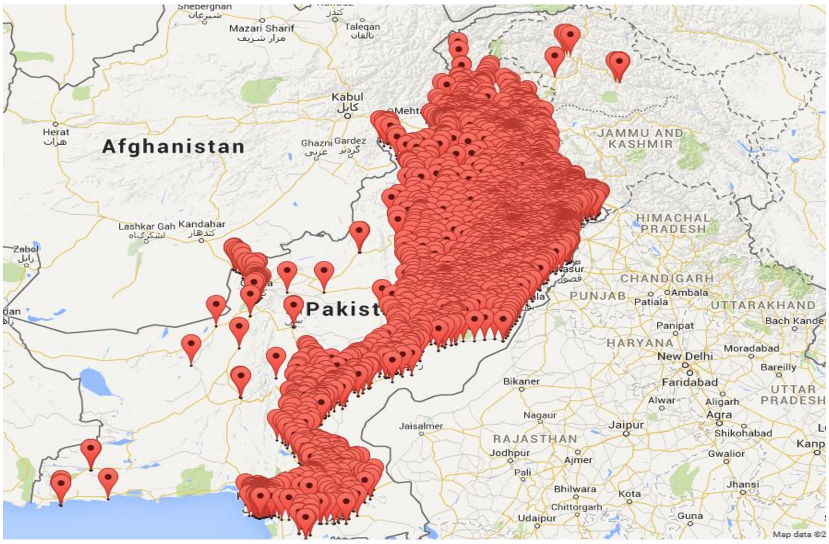


Fig 4.3 Display 6,000 users on Google Map before Clustering

Figures 4.1, 4.2, and 4.3 have shown the current geographical location and density of mobile users in different regions by plotting it on google map. The map shows that we have mobile users from all over country in dataset. However distributions of users are thickly or dense populated in different geographical locations of the whole country.

4.3.2. Single Level Clustering Results

Consider the figures above, where user clustering algorithm is not applied. Different number of mobile users scattered on the google map. These figures show that some users are located relatively close together in some regions, while other users are away from each other in some areas. In single level clustering, users are grouped into clusters according to their geographical vicinity. When a new user is registered, clustering algorithm will find a place in existing clusters within a defined geographical radius. If it fails to find a position in existing clusters, it will create a new cluster by adding new user. The number of users in different clusters is then displayed on map. The simulation results of single level user clustering on different datasets are shown in following figures:

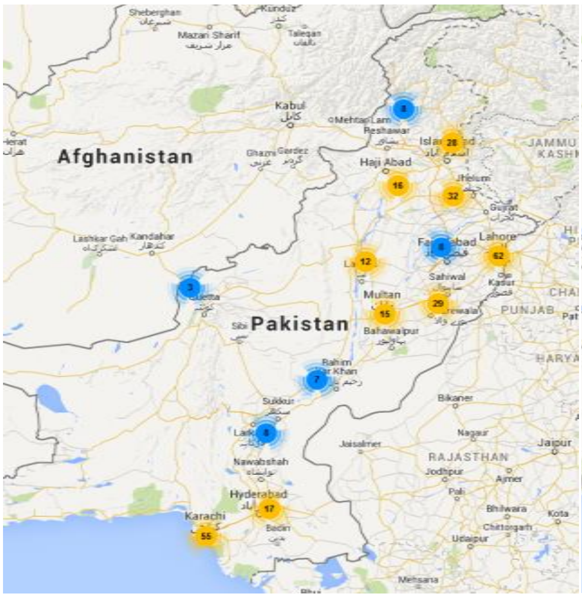


Fig 4.4 Single level clustered results of 500 users

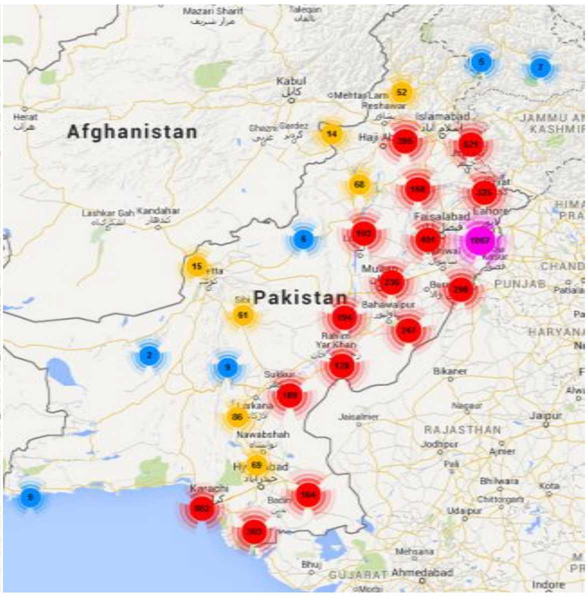


Fig 4.5 Single level clustered results of 6,000 Users

Figure 4.4 and 4.5 shows the complete single level user clustering results on google map. In these figures, clusters of different colors represent density of users in a particular geographical area.

4.3.3. Hierarchical Clustering Result

In our simulations, we used a total of four different datasets of mobile subscribers to evaluate the performance of clustering algorithm. The smallest of these datasets contained 100 mobile subscribers and the largest datasets contained 5000 mobile subscribers. In order to ensure diversity in our datasets, we randomly selected mobile subscribers from different geographical regions throughout the country. All the datasets with other clustering parameters used in our simulations are summarized in table 4.3.
To evaluate the performance of our proposed clustering algorithm, we implemented it on centralized application server using php programming language. The simulation results of each datasets are then shown on Google map using JavaScript and Google map API version 3. Our goals in conducting the simulation are as follows:

1. Evaluate the performance of the proposed algorithms on different datasets (different number of subscribers with other parameters) in wireless cellular network.
2. Study the effect of the threshold ***d*** (geographical radius) value on user clustering.

3. Study the effect of the Cluster Size on user clustering in different geographical regions.
4. Study the effect of the subscribers' device capability d_c on cluster head selection.

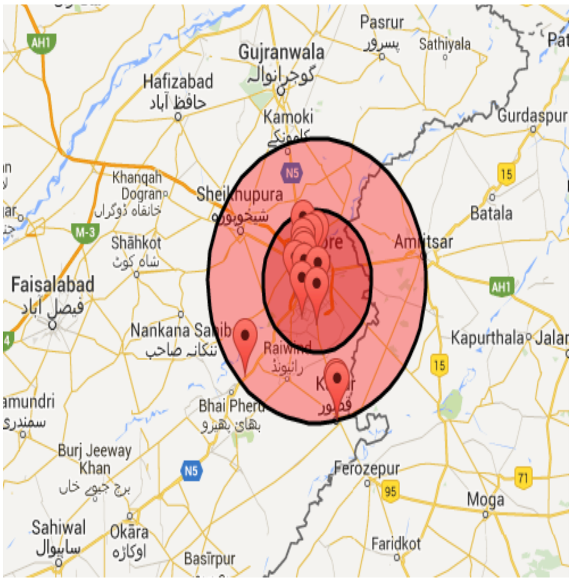
No of Users	Radius (Geo Distance)	Clusters	Sub-clusters
100	30 KM	23	2
	70 KM	17	6
	100 KM	14	6
500	30 KM	111	8
	70 KM	43	20
	100 KM	26	18
1000	30 KM	135	22
	70 KM	43	28
	100 KM	24	24
5000	30 KM	208	154
	70 KM	62	96
	100 KM	37	58

Table 4.3 Summary of results obtained from different simulations

We executed our proposed algorithms 3 times for each datasets and for each run we used different value of radius ***r*** (geographical distance). We used varying radius of 30KM, 70KM, and 100KM for each datasets. The results from these simulations against each datasets are summarized in table 4.3.

The simulation results on different datasets showed that the proposed algorithms for user clustering lead to better hierarchical clustering of mobile subscribers in cellular networks. It is also observed that most of the clusters and sub-clusters have been well distributed over different geographical regions as shown in figures 4.6 and 4.7.

The algorithm is tested for identifying the nearest subscribers to make clusters for different datasets. The simulation results revealed that the proposed algorithm can be used to identify the geographically nearest subscribers very effectively.



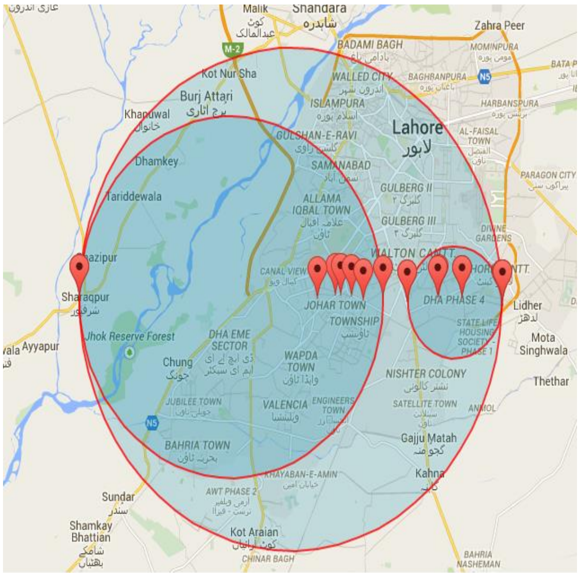


Fig 4.6 Hierarchical clustering Results on Google map

Fig 4.7 Cluster and Sub-clusters in a particular regions

The results of hierarchical clustering algorithm has been run 12 times with constant value of cluster size and varying different values for users and geographical radius for each run as shown in table 4.3. The performance of clustering algorithm for each run is shown in the following graph.

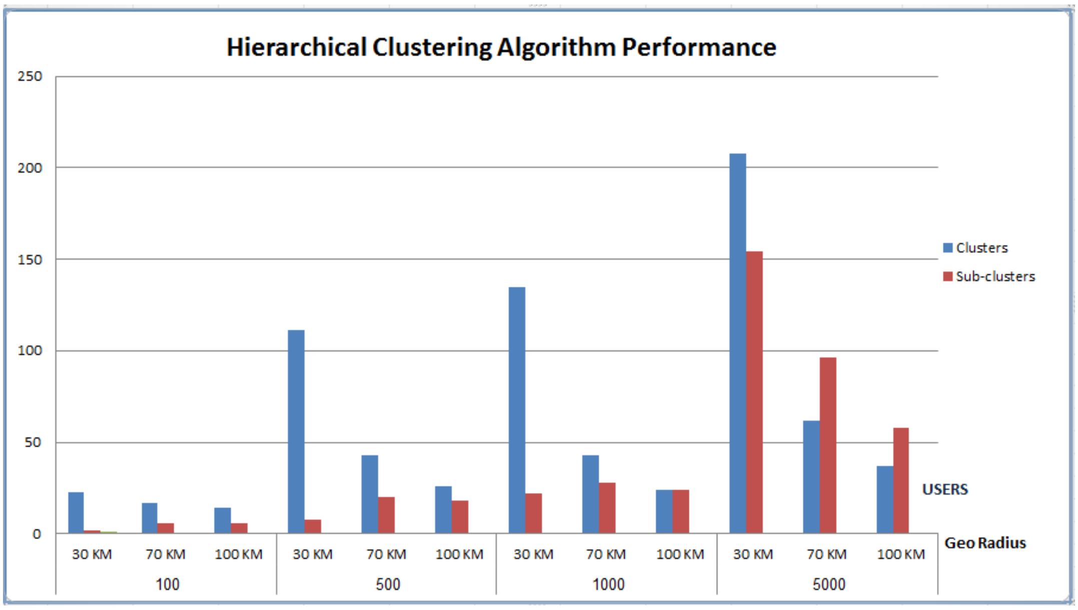


Figure 4.8 Hierarchical Clustering Performances

4.3.4. Prototype Functionalities

Below are the functionalities of our developed prototype on defined hardware platform that were achieved successfully by testing it in real time scenarios:

- Developed a client application which is successfully installed on different Android based mobile devices including Samsung Galaxy S3, S4, Sony Z1, QMobile A-50, A-8, A-150, X-350, and some other devices.
- A graphical user interface to operate the prototype was built. It ensured the desired results.
- The client application on subscriber device successfully determined its current location information from the cellular network.
- The mobile application was effectively gather subscriber preferences on initial application installation and then successfully sent subscriber preferences with their location information to server as a text message via SMS.
- The GSM modem in our scenario a Samsung Galaxy S3 mobile device which is connected to the computer system was successfully receive SMS from each new subscribers and further forward SMS contents to the centralized server.
- The prototype system succeeds in retrieve the SMS contents from modem and creates profile for each new subscriber.
- The developed prototype successfully grouped similar mobile subscribers into clusters based on their current location and also successfully selected head node for each new cluster.
- The prototype effectively notified every cluster head about its own status and other cluster members.
- The prototype system then successfully sent the requested information contents to the respective cluster head. The cluster head further sent to their cluster members and eventually each and every registered subscriber successfully received required data through their cluster head via SMS.

However, the current prototype for the proposed system was not fully developed. Though, having a live system on retrieving location-based information contents via SMS will prove that this prototype platform able to make the whole system design useful. Running this application live will bring benefits to mobile users where they can get their required information from internet regardless of any network connectivity no matter when and where they are.

4.4. Summary

In first section of this chapter, performance and efficiency of proposed algorithms is evaluated and validated using various parameters by simulating it on Google maps using Php, JavaScript, and Google maps API V3. The formations of user clusters and sub-clusters in different regions of Pakistan are shown in google maps. Simulation results on different data sets with differing features show that the proposed algorithm can form natural clusters in different regions according to the size of users. The simulation results shows that our proposed algorithms have better performance in terms of cluster formation, cluster head selection, strength of clusters or sub-clusters, stability of cluster/sub-cluster heads, and levels of cluster hierarchy. Simulation results on different data sets with varying parameters, it is observed that most of the clusters and sub-clusters have been well distributed over different geographical regions.

In second section, we implemented a prototype of the proposed system to validate our design and evaluate the performance of the complete system. Our evaluation results show that our proposed system is efficient, has low overheads, and is suitable for users' mobile devices.

Chapter 5
Conclusion and Future Work

5.1 Discussion

The key idea of our study was to deliver latest information contents to a group of mobile subscribers according to their current geographical location ubiquitously anywhere and anytime regardless of network connectivity. We used short message service as a data carrier for location-based message distribution. We assume that some of the subscribers' mobile devices can use as an intermediate processing nodes for user clustering and message distribution. The developed prototype system fairly confirms our assumptions.

5.2 Conclusion

In this study, an attempt has been made to design and develop a location-based, scalable, and geographically distributed system for mobile subscribers located in different geographical regions of a country. The developed system functionality is as follows: (1) in the very first step, mobile subscribers need to register for information services of their interest; (2) a clustering algorithm is performed to group users based on their geographical location; (3) algorithm to elects a head node for each new cluster is proposed; (4) Required information contents are transmitted to the cluster heads; (5) finally each cluster head receive messages through SMS from source (centralized server) and further forward to the remaining cluster members. And each targeted subscriber in the cluster eventually receives required information contents through its upper level cluster head. In order to location-based information delivery, the centralized server grouping subscribers into different clusters using their geographical location information. For this purpose, we have presented a novel hierarchical clustering algorithm for mobile users which consider the dynamic model of clusters. Our proposed algorithm can determine natural user clusters of different shapes and sizes in a particular geographical region because it is dynamically decides the merging and splitting of clusters in a particular region. A cluster head once elected, it stays as long as its message sending capability is within a predefined threshold value. Simulation results on different data sets with differing features show that the proposed algorithm can form natural clusters in different regions.

However, currently we have developed a prototype of the proposed system to deliver messages according to subscribers' geographical location. Research as well as development work is continuing in this project with some additional features being added to the system. The current prototype allows a mobile subscriber to send request for service subscription and after clustering subscriber receive their required information on any Android compatible mobile phone via SMS. The contents of received message can be viewed on subscriber mobile phone.

5.3 Future Work

Though current prototype can be extended in the following directions in future:

5.3.1. Diverse Platform Compatibility

Although the current developed prototype is only compatible for any Android-based mobile phone. In future it can easily be designed in such a way that it can be work for any type of GSM compatible mobile phone.

With the advent of time, technology is advancing day by day. Our prototype system will further be developed and enhanced. Currently the developed prototype is available for Android-based mobile devices. Future work comprises development of the application for four major platforms: iPhone, Windows, Blackberry, and Symbian.

5.3.2. Interest-Based Clustering

We have implemented and simulated our proposed clustering algorithm based on subscribers' geographical location with users default interest. The future work will be primarily focused on user clustering based on multiple users' interests and level of interest with their geographical vicinity. In future we will group users based on their interest so that any user may be a member of more than one cluster in the meantime and receives messages from different cluster heads according to their cluster membership. But can be a head node for only single cluster or sub-cluster at a time.

However, more effort requires to be done to realize and validate our proposed system. Our future work includes the implementation of the proposed distributed architecture using as starting point the current developed prototype of the service.

5.3.3. Compatible for Diverse Networks

Although the current developed prototype is only compatible for any Android based mobile phone. In future it can be designed in such a way that it can be compatible for any type of mobile phone device. Also the prototype version of the system is currently implemented on the existing GSM communication network so in future it can easily be designed and configured for use on any other kind of digital cellular communication network that provides text messaging functionality such as personal handy phone system (PHS) or personal access system.

5.3.4. Development of Live Software Application

This research emphasis on the techniques to retrieve latest information contents on mobile devices from internet via short message service (SMS). No actual or live application was developed. Therefore, for future research live application software should be developed in order to verify that our developed prototype technique is capable to deliver updated information contents to various mobile subscribers regardless of any network connectivity. Our future work also can be made on analyzing the short messaging cost, which is more practical, inexpensive, and a useful service with less limitation features.

References

[1] S. Kumar, M. a. Qadeer, and a. Gupta, “Location based services using Android (LBSOID),” *Internet Multimed. Serv. Archit. Appl. (IMSAA), 2009 IEEE Int. Conf.*, 2009.

[2] D. Mohapatra and S. B. Suma, “Survey of location based wireless services,” *2005 IEEE Int. Conf. Pers. Wirel. Commun. 2005. ICPWC 2005.*, pp. 358–362, 2005.

[3] P. Bellavista, A. Küpper, and S. Helal, “Location-based services: Back to the future,” *IEEE Pervasive Comput.*, vol. 7, no. 2, pp. 85–89, 2008.

[4] Y. Shang, W. Zeng, D. K. Ho, D. Wang, Q. Wang, Y. Wang, T. Zhuang, A. Lobzhanidze, and L. Rui, “Nest: Networked smartphones for target localization,” *2012 IEEE Consum. Commun. Netw. Conf. CCNC’2012*, pp. 732–736, 2012.

[5] V. Roto, R. Geisler, A. Kaikkonen, A. Popescu, and E. Vartiainen, “Data Traffic Costs and Mobile Browsing User Experience,” *Proc. 4th MobEA Work. Empower. Mob. Web*, vol. 3, pp. 1–6, 2006.

[6] J. Codr, “Wireless Options for Providing Internet Services to Rural America,” 2008. [Online]. Available: http://www.cse.wustl.edu/~jain/cse574-08/ftp/rural/#wimax.

[7] Z. Jerzak and C. Fetzer, “Handling overload in publish/subscribe systems,” in *Proceedings - International Conference on Distributed Computing Systems*, 2006, no. July 2015.

[8] C. Liu, Y. Liu, X. Ma, and J. Gao, “An application scheme of publish/subscribe system over clustering mobile ad hoc networks,” *2010 6th Int. Conf. Wirel. Commun. Netw. Mob. Comput. WiCOM 2010*, pp. 3–6, 2010.

[9] C. Peersman, S. Cvetkovic, P. Griffiths, and H. Spear, *The Global System for Mobile Communications Short Message Service*, vol. 7, no. 3. 2010, pp. 15–23.

[10] S. A. Baset and H. G. Schulzrinne, “An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol,” in *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, 2006, pp. 1–11.

[11] D. Friel and L. Kilmartin, “An automated stock price delivery system based on the GSM short message service,” *ICC “98. 1998 IEEE Int. Conf. Commun. Conf. Rec. Affil. with SUPERCOMM”98 (Cat. No.98CH36220)*, vol. 3, pp. 1591–1595, 1998.

[12] R. Pries, T. Hobfeld, and P. Tran-Gia, “On the Suitability of the Short Message Service for Emergency Warning Systems,” *2006 IEEE 63rd Veh. Technol. Conf.*, vol. 2, no. c, pp. 991–995, 2006.

[13] S. P. Singh and S. C. Sharma, “A Survey on Cluster Based Routing Protocols in Wireless Sensor Networks,” *Procedia Comput. Sci.*, vol. 45, pp. 687–695, 2015.

[14] Y. Liu, S. Shi, and X. Zhang, “Balance-Aware Energy-Efficient Geographic Routing for Wireless Sensor Networks,” *2012 8th Int. Conf. Wirel. Commun. Netw. Mob. Comput.*, pp. 1–4, 2012.

[15] A. Dumitru, A. Marinescu, C. Dobre, and V. Cristea, “Routing Protocol for Urban Mobile Networks based on Geographical Location,” *IEEE Int. Conf. Intell. Comput. Commun. Process.*, pp. 479–482, 2010.

[16] C. Truong and K. Romer, “Efficient geocasting to multiple regions in large-scale wireless sensor networks,” in *Proceedings - Conference on Local Computer Networks, LCN*, 2012, pp. 453–461.

[17] C.-T. C. C.-T. Chen, S. Tekinay, and S. Papavassiliou, “Geocasting in cellular ad hoc augmented networks,” *2003 IEEE 58th Veh. Technol. Conf. VTC 2003-Fall (IEEE Cat. No.03CH37484)*, vol. 3, pp. 1858–1862, 2003.

[18] R. Kruthika, S. S. B, and G. Harish, “Geocasting and Multicasting Routing Operation in Mobile Ad Hoc Network,” no. 05, pp. 135–139, 2013.

[19] C. Maihofer, “A survey of geocast routing protocols,” *IEEE Commun. Surv. Tutorials*, vol. 6, no. 2, pp. 32–42, 2004.

[20] T. S. Kote, S. R. Jeyashankher, and L. Chandran-Wadia, “A large scale publish-subscribe platform for information delivery to mobile phones,” in *3rd IEEE/Create-Net International Conference on Communication System Software and Middleware, COMSWARE*, 2007, pp. 387–394.

[21] A. Hussein, S. Yousef, and O. Arabiyat, “A Load-Balancing and Weighted Clustering Algorithm in Mobile Ad-Hoc Network,” *kaspersky.co.za*.

[22] Y. Zhao, X. Wang, U. Heo, and J. Choi, “An efficient tree-based geocast routing algorithm for connecting heterogeneous networks,” *Proc. - 2010 Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discov. CyberC 2010*, pp. 331–338, 2010.

[23] L. Junhai, Y. Danxia, X. Liu, and F. Mingyu, “A survey of multicast routing protocols for mobile Ad-Hoc networks,” *IEEE Commun. Surv. Tutorials*, vol. 11, no. 1, pp. 78–91, 2009.

[24] Q. He, “A review of clustering algorithms as applied in IR,” *Grad. Sch. Libr. İnformation Sci. …*, vol. 2, no. 11, pp. 1–33, 1999.

[25] S. a Baset and H. G. Schulzrinne, “An Analysis of the Skype Peer to Peer Internet Telephony Protocol,” pp. 1–11, 2006.

[26] A. P. Singh and N. Sharma, “The Comparative Study Of Hierarchical Or Cluster Based Routing Protocol For Wireless Sensor Network,” vol. 2, no. 6, pp. 1806–1813, 2013.

[27] A. A. Abbasi and M. Younis, “A survey on clustering algorithms for wireless sensor networks,” *Comput. Commun.*, vol. 30, no. 14–15, pp. 2826–2841, 2007.

[28] http://www.pta.gov.pk/annual report 2013-2014.

[29] http://www.longitudestore.com/haversine-formula.html.