A Fine-Grained Addressing Concept for GeoCast

Peter Coschurba and Kurt Rothermel

Institute of Parallel and Distributed High-Performance Systems (IPVR),
University of Stuttgart,
Breitwiesenstraße 20-22, 70565 Stuttgart, Germany
E-mail: Peter.Coschurba@informatik.uni-stuttgart.de
Phone: +49 711 7816 416, Fax: +49 711 7816 424

Abstract

GeoCast provides the functionality of sending messages to everyone in a specific area. So far, only the addressing of larger two-dimensional areas was possible. For the use in an urban environment it is crucial that small and three-dimensional areas can be addressed. For example, GeoCast can then be used to send lecture notes to all in a classroom. In this paper we describe a fine-grained addressing concept for GeoCast that supports such areas. In addition we present an architecture that allows the use of that addressing concept together with the GeoRouting-approach developed by Navas and Imielinski [8]. We also present some modifications necessary to enhance the scalability of GeoCast.

1 Introduction

If we look at the means of communication that human beings use, we find that there are different categories of how we determine the receiver of a message. If we speak, everyone who is at the same location will get the message. We can adjust the area in which the message is perceived by changing the tone of our voice. If we shout, the area gets larger and if we whisper the area becomes smaller. If we write a letter, we address someone specific and only that person will get the message. Finally, if we publish something, then everyone who is interested in it can get that information. These ways of communication are also reflected in the Internet. We can send messages to a specific receiver using unicast. If we use multicast, everyone who is interested in the information can tune into the group and receive the messages. Finally, as an equivalent to speech, which is a location specific way of communication we find GeoCast [1, 2].

GeoCast is a relatively new means of communication. A sender specifies the area in which the message should be received by everyone. The receivers do not need to be known and do not have to decide that they want to receive the messages. They get the message when they are in the area that the sender specified and they do not get it if they are outside that area. Because many messages are not only interesting for users who are currently in the area, but also for those who arrive some time later, the concept of lifetime was introduced to GeoCast. The lifetime of a message determines how long this message will be available in the target area. It can be compared to a poster. While the poster is up, the message can be read by everyone close enough. Messages with a lifetime can be used to warn people of danger, to announce new services or for many other purposes.

In order to use such means of communication we find that the receivers have to have devices that are capable of determining their current position, maintaining a connection to the Internet and doing the necessary computation. Devices, such as future mobile phones or PDA's will become artefacts of our daily life.

GeoCast is still a fairly new concept, so it is at the very beginning of its evolution. One of the main issues in GeoCast is how the area, that defines the receivers of the message , can be specified. GeoCast has been mainly driven by the work of Julio Navas and Thomasz Imielinski. Their protocols are optimized for larger two-dimensional areas [3]. Therefore we call this kind of addressing coarse-grained. Coarse-grained addresses can be used to send a message to all people in a wood, that a wildfire is happening or send a message to a part of a city announcing that a new restaurant has opened. Considering the needs that arise if GeoCast is to be used in an urban environment we find that having coarse-grained addressing is not sufficient [4].

Scenarios that show possible usages of what we call fine-grained addressing are to send lecture notes or subtitles for the hearing impaired to everyone listening to a talk, to distribute some information on a floor (e.g. broken lift) or to send a message to everyone on a bridge (e.g. warning of a traffic jam). Common to these scenarios is that they rely on the third dimension and that the addressed areas are small.

An important aspect of addressing a geographic area is how location information can be expressed. There are two fundamental possibilities. One can express the location of a person by stating his coordinates in a predefined coordinate system like WGS84 [5]. Or the position information can be expressed using a symbolic identifier that states in what area a person is. Whereas coordinates are dominant in the outdoor use (e.g. used by GPS [6]) symbolic identifiers are dominant in the

indoor use (e.g. used by Active Badge System [7]). To make GeoCast available in all kind of areas, it has to work together with both kinds of position information and therefore with all kinds of positioning systems.

In this paper we will look at the technology that is used for GeoCast. We will extract the requirements that GeoCast has to meet in order to be usable in urban as well as in rural areas. A fine-grained addressing concept will be presented that meets the requirements and finally we will show how this addressing concept can be implemented using the GeoRouting presented by Navas and Imielinski. The local dissemination of GeoCast messages imposes scalabilty problems. We will introduce a notify-pull mechanism, the so-called message announcements, in order to make GeoCast more scalable.

This paper is organized as follows: In the next section we will look at the related work. Section three deals with the requirements for a more sophisticated fine-grained addressing concept. In the following section we will present an addressing concept that meets these requirements. In section five we will look at the realisation and implementation of the addressing concept. And finally, we will conclude our paper with an outlook on future work.

2 System Model

In this section we will have a closer look at the technology that forms the foundation of this paper. We will look at the current state of GeoCast, at the way how addresses can be constructed today and define some terms that are used throughout this paper.

The focus of Navas and Imielinski is the message forwarding. They have looked at different means for transporting the message towards the target area and have defined a protocol called GeoRouting that is able to route geographic messages with acceptable performance [8, 9]. Such a GeoCast using GeoRouting is shown in Figure 2-1. The sender S sends a message to the grey area. The node E is inside the area and therefore should receive the message. To develop an addressing concept was not their primary concern and thus their concept of defining target areas is fairly simple. The **target area** of a message is the geographic area in which the message is to be distributed. In other words, the target area is the area that the sender specifies and in which every participating node will receive the message.The target area can be described using a polygon of the area-boundary, specified through the coordinates of the edges in WGS 84 coordinates [1]. Because only two-dimensional coordinates have been chosen, it is not possible to address a three-dimensional area. The whole protocol is optimized for larger target areas, which normally span several networks.

In order to perform GeoCast several software components are needed:

The **GeoHost** is the software running on the client. It is responsible for sending and receiving the messages on the device. In order to send a message, the GeoHost builds packets and forwards them to the next GeoRouter. Messages that are to be received are broadcasted by the GeoNodes on the networks in the target area. Because there might be devices connected to the network that lie outside the target area, it is necessary that the client itself checks whether it is actually supposed to receive the message or not. If the GeoHost is inside the target area, the message is forwarded to the appropriate application. Therefore, in Figure 2-1 GH1 accepts the message and GH2 discards it.

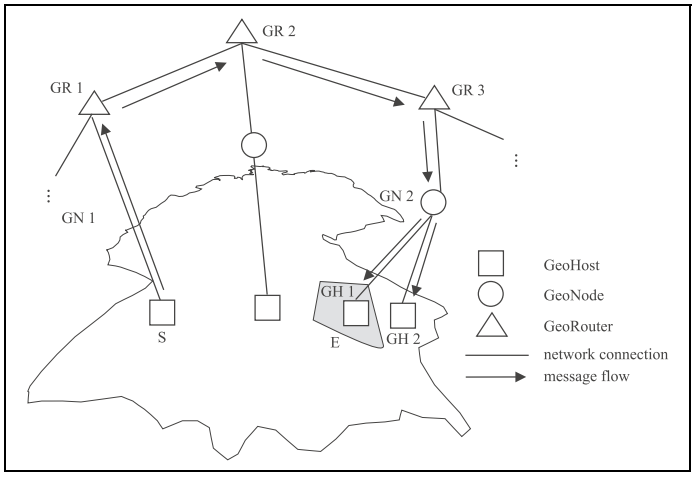

Figure 2-1. GeoCast using GeoRouting

The **GeoNode** is a special node that is responsible for distributing the message in the target area. In addition, it stores messages which have a lifetime. The messages are distributed by the GeoNode until the lifetime expires. There should be one GeoNode for each IP-subnetwork. A GeoNode services the network that it is directly connected to. The geographic area that is covered by that network is called the **service area** of the appropriate GeoNode.

Another component is the **GeoRouter**. Each GeoRouter knows its service area. The service area of a GeoRouter is the aggregate of the service areas of the GeoNodes that are assigned to it. Normally these GeoNodes are directly connected to the GeoRouter, but there is also the possibility to connect a GeoNode using tunnelling. The GeoRouters use the GeoRIP protocol to exchange the information about their service areas. This results in every GeoRouter knowing the service areas of all other GeoRouters. If a GeoCast message arrives at a GeoRouter it checks which service areas of the other GeoRouters overlap with the target area of the message and forwards the message towards them. If the target area overlaps with its own service area, then the message is forwarded to the appropriate GeoNodes.

3 Requirements on Fine-Grained Addresses

The main requirement for the fine-grained addressing concept is the ability to address small target areas. Additionally, three-dimensional target areas should be addressable in order to allow target areas like rooms. The lack of such a concept is the major flaw of existing GeoCast systems that prevents the use in an urban environment.

For GeoCast to become a success it must be efficient and scalable. So the addressing concept should not only take into account what the user might want to address, but also if addressed areas can be processed efficiently. As we have seen, there are two kinds of operations necessary in order to deliver a message. The GeoRouters check whether the target area overlaps with some service areas of other GeoRouters and the GeoHosts calculate if their position is within the target area. Especially the first check is important, because for routers it is critical to reach the forwarding decision fast, in order to prevent congestion. But also the check on the client can be critical if we want computationally "weak" devices to participate. Another issue is bandwidth consumption. As the specification of a three-dimensional body can use a large amount of space, it is an important requirement that the space used by the address stays small, in order to increase the amount of payload that can be transported in each packet.

The major strength of the existing protocol is the message forwarding itself. GeoRouting offers a good means to transport messages to two-dimensional areas. If possible, the fine grained-addressing concept should work together with the GeoRouting of Navas and Imielinski.

Hence the requirements are:

- Support for efficient handling of small target areas
- Ability to address three-dimensional target areas
- Efficient overlap calculation (for the GeoRouters)
- Efficient include operation (for the GeoHost)
- Compact addresses
- Interoperability with GeoRouting

4 Fine-Grained Addressing Concept

In this section we will present our addressing concept. As stated above, two addressing methods have to be considered. On the one hand there are coordinates and on the other hand there are symbolic identifiers. Both describe location using different concepts. The first denotes a position whereas the latter states in what area somebody is. It is important to note that there is normally only one kind of position information available in an area. Either the sensor systems in the infrastructure delivers coordinates or symbolic identifiers. To have both (expensive) infrastructures in the same building is only a rare case restricted to a few research institutions. So we can assume that inside such an area there is only one kind of position information available. Our fine-grained addressing scheme consists of two kinds of addresses. One kind is based on coordinates and the other is based on symbolic identifiers. Depending on the technology used in the target area the appropriate address can be used.

4.1 Geographic Coordinate Based Addresses

For coordinate based addresses we have to differentiate between two-dimensional and three-dimensional target areas. Two-dimensional ones can be specified using the concepts of the existing protocols. They allow to address polygons and circles. If 8 Bytes represent an element of a coordinate then the accuracy of the coordinates is below a millimetre, which should be enough for all kinds of applications. In the case of the polygon, one specifies all the edges using two-dimensional WGS84 coordinates. For the circle the centre and the radius have to be specified.

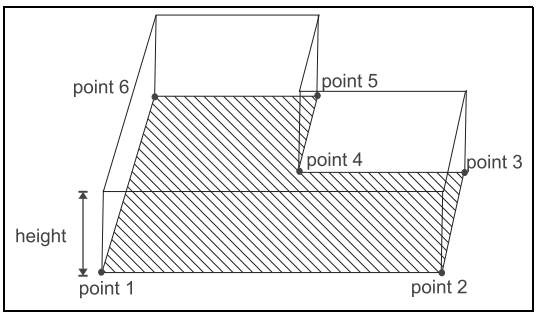

Figure 4-1. 2.5-Dimensional Target Area

For three-dimensional addresses the situation is more complex. On the one hand we want to have small addresses that can be processed efficiently and on the other hand we want to give the user the possibility to address all target areas that he might need or want to address. If we look at what that target areas might be, we find that they can be modelled by using so called 2.5-dimensional bodies. Targets like a room, a floor or a house have in common that they can be defined by specifying the base and adding a fix height (see Figure 4-1). An example is shown in picture where a room is addressed.

Such addresses have the advantage that they are compact, using only slightly more space then two-dimensional ones (8 Bytes for each coordinate dimension and another 8 Bytes for the height) and all the overlap and inclusion calculations can be done using

the same algorithms used for two-dimensional figures adding only a check for the height. In order to address all kind of targets we provide two base figures. One is a polygon with height and the other is a cylinder. An example for such and address would be the first floor of the computer science building of the university of Stuttgart:

```
3D_Polygon((48.431782N; 9.7410E; 508) (48.43169N; 9.7398E; 508)
(48.431815N; 9.7398E; 508) (48.43195N; 9.7363E, 508) 2.6)
```

The coordinates of the edges are listed counter-clockwise and the height is given in meters. The third parameter of the coordinates is the height over sea level.

4.2 Symbolic Identifier Based addresses

Symbolic identifiers describe an area in which the object is located. Such an area is in most cases a room[1] or an open area with a distinct extension. One could think that providing that identifier would be enough to address the room. But unfortunately these identifiers are only unique within a limited area. The addressing scheme used by company A can also be used by the neighbouring company B. So there could be two rooms with exactly the same identifier. But in order to use the symbolic identifiers as an address they have to be globally unique. To achieve that, we add to each symbolic identifier the area in which the ID is unique, that is, the area that is covered by the organisation to which the room belongs. This area can be a large two-dimensional area, in case of a large company or it can be only the floor of a large building when we look at a small company or a local branch of some company. So an address that is based on symbolic identifiers consists of two parts, the geographic address denoting the area in which the identifier scheme is valid and then the identifier itself.

Many symbolic identifier schemes are built hierarchically, thus implementing the “inside” relationship. The identifier reflects in what building and in what floor the room is. So, for example the symbolic identifier “22.2.018” in the Computer Science Dept. of Stuttgart implicitly states that the room denoted by this symbolic identifier is located in building 22 on the second floor. For GeoCast this is - of course - a nice feature. Besides single rooms, it allows to address all rooms on a specific floor or even in a building. In order to use the knowledge inside the identifiers, there must be a understanding of the semantics of the symbolic identifier. Therefore it is necessary to have a common format, how these identifiers are built. We use a scheme that will be met by most local schemes:

“`BUILDING.FLOOR.ROOM`”.

If another scheme is used, there has to be a mapping to this scheme. This mapping function has to be provided to the clients in order to make use of the knowledge embedded into the identifier.

To make things more clear we will look at our example. We want to address room 18 on the second floor of building 20 at the Computer Science Dept. of Stuttgart. The appropriate address looks as follows:

```
ID((2D_Polygon((48,431782N; 9,7410E) (48,43169N; 9,7398E) (48,431815N;
9,7398E) (48,43195N; 9,7363E))(20.2.18))
```

The first part describes the area in which the identifier scheme is valid, or in other words which is encompassed by the model. The second part is the identifier itself. If we wanted to address the whole floor we would just omit the 18 at the end of the identifier.

5 Realisation

5.1 Architecture

To integrate fine-grained addresses with the GeoRouting of Navas and Imielinski is the main focus of this paper. In order to achieve that, at least the GeoHost has to be modified. But in order to achieve a good performance and to get a scalable solution changes at the GeoNode are also needed. The basic architecture remains the same. GeoHosts have to be present on every node that wants to send or receive messages. A GeoNode is needed for each participating subnetwork and there are GeoRouters, which do not have to be modified.

However, the interface between the application and the GeoHost and between the GeoNode and the GeoHost have to be changed. The interface of the GeoHost that an application can use to send a message is modified to support the new types for target areas. The function is `sendGeoMessage(Area, Port, Lifetime, MessageID)`. Thus, the application can specify the following parameters:

- Area: The area information consists of an area type and the parameters (points and height or radius information) that specify the target area. Each point is provided using 8 bytes for longitude, latitude and, in case of a three-dimensional area, for the height information. For areas whose addresses are built using an area and a symbolic identifier, this can be specified too. The type and parameters are:

1. Caused by the fact that most sensor systems use technology that can not penetrate walls.

    - 2D_Polygon: For a polygon all points that specify the edges are listed counter clockwise.
    - 2D_Circle: For the 2D_Circle the centre point is specified and then the radius is provided in metres.
    - 3D_Polygon: A list of three-dimensional points and the height of the figure is specified. The height is given in metres above the base of the figure.
    - 3D_Circle: A cylinder is specified by specifying the point that is the centre of the circle that forms the base. The radius and the height are also provided in metres.
    - ID: An identifier based address always consists of two parts. First an area is provided and then the identifier. (see Section 4.2)
        - Area: Any of the areas described above
        - ID: A string that contains the identifier. Points are used as delimiters.
- Port: The port specifies which applications will receive the message.
- Lifetime: The lifetime specifies for how many seconds the message will be distributed by the GeoNodes. The time starts when the GeoNode receives the message.
- Message ID: The sender creates a unique message identifier for each message.

As stated above, GeoRouting is used to forward the messages. It offers an acceptable performance and has already implemented by Navas and Imielinski. But GeoRouting operates only on two-dimensional target areas. So it seems that the GeoRouter have to be modified to route on three-dimensional target areas. But if we look at the requirements, we see that the efficiency at the forwarding decision is one of the key requirements. The routing on three-dimensional target areas would make the forwarding decision more complex. Therefore we have chosen another alternative. We will do the forwarding decision based on two-dimensional areas. For this we introduce the concept of the **scope** of a message.

We distinguish between the address of a message, which denotes the target area and the scope, which is a simplified target area used only for routing. In order to allow an efficient routing, we use the scope, a two-dimensional area, to forward the message. All GeoNodes, that are inside that scope will get the message. They will then look inside the message and evaluate the address. If their service area overlaps with the target area, the message will be processed like normal, if not it will be discarded. So messages are only distributed in networks that service the target area. This modification in the GeoNodes makes it possible to have an efficient routing and to save the bandwidth in the networks that overlap with the scope but not with the target area.

But how is the scope generated? In case of a two-dimensional target area, the scope is just identical to the target area. If a three-dimensional target area is specified, then the third-dimension is omitted. For addresses based on symbolic identifiers, the geographical part is used to generate the scope. The result of all these operations is a two-dimensional area that can be used by GeoRouting. This is used to build a header conforming to the GeoRouting. As a result we have two headers, the outer header is conform to the GeoRouting and specifies the coarse-grained scope of the message and an inner header, that specifies the fine-grained target area.

To clarify the architecture, we will look at what happens when a GeoCast message is sent:

1. The sender specifies a target area. This is done using the fine-grained addressing concept. The message is then transferred by the sending application to the GeoHost.
2. The GeoHost calculates the scope. It then creates a message as it can be used by the GeoRouting and encapsulates the original message with the original header in it. The header for such fine grained GeoCast message is shown in Figure 5-1.This message is then processed as before the modification. It is forwarded to the next GeoRouter and from there to all the GeoNodes in the scope.
3. The GeoNode checks whether its service area overlaps with the target area and forwards the message to the network if it does.
4. The GeoHost on the receiving device checks whether the device is actually in the target area. This is done by comparing the position information with the address information in the message header. In case of an identifier based address, the device checks first if the area of the address equals the validation area of its identifier based scheme. In a second step it checks whether the identifier that describes its location is either equal or enclosed in the identifier of the address. If the position is inside the area, the message is accepted and forwarded to the appropriate application, if not it is discarded.

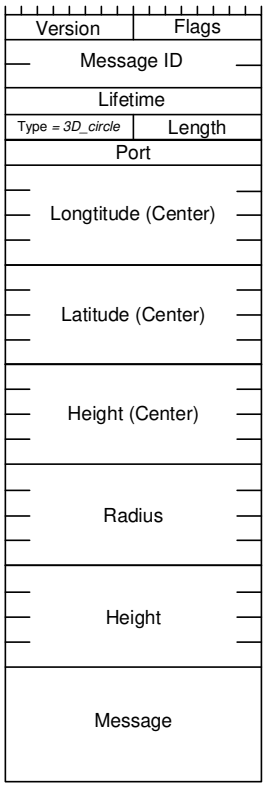

Figure 5-1. Message Header for Coordinate Based Target Area

5.2 Message Announcements

So far we have shown how to send messages using the fine grained addressing concept. But if we look at the lifetime of a message a new problem arises. Originally, the target areas were rather large, spanning several networks. In addition it was assumed that only few messages are sent using GeoCast. With the new fine-grained addresses a different situation arises. First of all, the target areas get smaller. The average target size will probably be a room. Because most networks span more then only one room, there might be several target areas overlapping with the service area of the network.

So far, the lifetime was realized, by just re-sending the messages during the lifetime. Clearly, this is not scalable if we have lots of messages. Many of these messages would be sent without anyone being interested in it. Ideally, the messages would only be resent when someone enters the target area. Because the receivers do not have to announce their position and presence, this is not possible. The idea is to send only short messages that include the address of the message and the message identifier. The GeoHost on the mobile device checks, if it is in the target area and if it already received the message. If it determines that it is inside the target area, but has not received the message so far, it issues a message request, requesting the message to be resent. Upon receiving such a request the GeoNode will resend the whole message.

Whereas previously the whole messages had to be resent during the lifetime, now only short announcements are sent. This results in a large decrease of bandwidth that is used for the local dissemination of the GeoCast messages and thus makes GeoCast more scalable. Especially, if we take into account that the smaller areas will lead to an increase in message size. There is rarely a usage scenario for large areas where one wants to send, for example, a power point presentation. But to send the slides during a presentation to all people in room makes sense.

Such a message announcement consists of two parts:

- The unique message ID: Every message gets a unique message identifier. This identifier is constructed by the sender.
- The target area: Each message announcement contains the target area that forms the address of the message. The format is exactly the same as in the message header.
- The port: This allows the potential receiver to determine if there is an application registered for that port, which is interested in such messages. Only those messages are requested for which receiving applications are registered which eliminates unneccessary transmissions.

Since the size that is needed to store the target area is not fixed, we can not state how large such an message announcement will be. But if we assume that the standard case will be to address a room, then such a message announcement will use 140 Bytes. So even in the unlikely case that 500 messages have a target area that overlaps with the service area and have a valid lifetime, that is a lifetime that has not yet been expired, only about 6% of the bandwidth of a 802.11b network would be used to send the announcements every second. If we assume that the average message is about 10 kilobyte, it would take about 6 seconds while using up all the bandwidth to resend the messages. So clearly we enhance the scalability of GeoCast by using message announcements.

The rate at which the messages are repeated is subject to parametrization. If the announcements are sent more frequently, they use up more of the bandwidth but people entering a target area will get the message earlier. If the messages are sent less frequently, bandwidth is saved, but the receivers will get the message later.

We think that four seconds between announcements is a good value for networks that have a bandwidth of about 10 megabit per second (like 802.11b). Four seconds is a rather small time for human beings, so the user will not notice a large delay. On the other hand, the consumed bandwidth is quite low. In the above example, we could send 500 message announcements by using only about 1.5% of the available bandwidth. Again, if we sent the whole messages all the time, we would use about the whole bandwidth in order to resend the messages all four seconds. Clearly this is not possible. Of course the frequency parameter is subject to local optimization. It has to be adjusted to the local needs and depends on the average number of messages and the bandwidth of the network.

In order to understand how this protocol works we will look at what happens when a message with a valid lifetime arrives at the GeoNode:

1. Upon reception of a new message the GeoNode checks if its service area overlaps with the target area. If it does the message is sent.
2. If the message has a valid lifetime then the message is stored, and the target area is used to build a message announcement. Every four seconds the announcement is sent (together with the other announcements) until its lifetime expires. Before the announcements are sent, the GeoNode checks if the lifetime has not expired. If it has expired, then the message is just discarded.
3. A GeoHost that receives the announcement uses the message Id to check whether the message was already received earlier. If not, it checks if its position is inside the target area. If it is inside the target area and has not received the message yet, it sends a message request to the GeoNode. Such a message request is represented by the message id.
4. When the GeoNode receives a message request, it re-sends the message on an broadcast channel, so that other devices that have just entered the area do also get the message.

6 Conclusion

The advent of location-aware applications and ubiquitous computing has created a need for geographic communication. The existing approach for GeoCast, however, has some disadvantages that have prevented a wide spread use. So far one of this disadvantages is the lack of a sophisticated addressing concept that allows the addressing of the intended areas. Another problem has been the scalability issue, i.e. if lot's of messages arrive for several closely connected areas. For both problems we have presented an solution.

Still more problems remain that need to be solved before GeoCast will become part of our everyday communication. So far, there are no real applications that make use of GeoCast. Without such applications, which provide a clear benefit to the user, GeoCast will not be used. We have shown several usage areas where GeoCast makes sense, so we assume that these applications will arrive soon. Actually, we have developed an application that allows the sending and receiving of geographic e-mails and messages. GeoRouting is relatively efficient, but still several magnitutes worse then all other routing methods for unicast or multicast. This remains an area of research. If GeoCast is to become popular, the message forwarding mechanism has to become more scalable and more efficient.

However, the foundations are laid, and GeoCast will become more and more popular. It's use will grow together with the number of location-aware applications. And therefore we think that it is worthwhile to do research in order to solve the above problems and create a means of communication that is beneficial to it's users.

7 References

[1] T. Imielinski and J. C. Navas, “GPS-Based Addressing and Routing” IETF RFC 2009, November 1996

[2] Young-Bae Ko and Nitin H. Vaidya, “Geocasting in Mobile Ad Hoc Networks: Location-Based Multicast Algorithms”, Proceedings of the 2nd Workshop on Mobile Computing Systems and Applications (WMCSA '99), 1999

[3] T. Imielinski, J. C. Navas, “GPS-Based Geographic Addressing, Routing and Resource Discovery”, *Comunications of the ACM Vol. 42*, April 1999

[4] Julio C. Navas. "Geographic Routing in a Datagram Internetwork". Ph.D. Thesis. Computer Science. Rutgers, The State University of New Jersey. May 2001

[5] WGS84 Military Standard Document available at http://164.214.2.59/publications/specs/printed/WGS84/wgs84.html

[6] Global Positioning System Overview at http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html

[7] R. Want, A. Hopper, V. Falcao, J. Gibbons, “The Active Badge Location System” ACM Transactions on Information Systems, Vol. 10, No. 1, January 1992, pp 91-102

[8] J. C. Navas and T. Imielinski, "Multi-hop Dynamic Geographic Routing", Technical Report (LCSR-TR-364), Rutgers University Computer Science, May 29, 1998

[9] Julio C. Navas and Tomasz Imielinski, "On reducing the computational cost of Geographic Routing" Technical Report (DCS-TR-408). Rutgers University Computer Science. January 24, 2000.