Chapter 1

Introduction

The present scenario is in the era of communication or the age of pervasive / ubiquitous computing. Identification of any person or system at any given location at any given time based on global positioning systems is expected to be the norm. This proliferation of connectivity can be understood by the fact that already the connected systems on the network are capable of reaching nearly twice the population of the earth, even when half the world is not connected. Users expect to be no longer constrained by slow connections, limited speed of computer processors and high costs of storage, for inclusive technology like cloud computing has evolved, thus offering the unconnected world an inexpensive option to choose the best of features provided by the Information and Communication Technology (ICT).

Researchers are already doing some rethinking on the internet underlying architecture, and are even considering the replacement of the networking equipment and rewriting software to meet the growing demand. One challenge in any reconstruction, though, will be balancing the interests of various constituencies. Recently the national science foundation (NSF) USA has proposed to build an experimental research network known as Global Environment for Network Innovations, or GENI. Several projects are being funded in universities and else where through a program called Future Internet Network Design (FIND).
The European Union has also backed research on such initiatives, through the program known as Future Internet Research and Experimentation (FIRE).

A new network or an existing network with enhanced capability is a need (needs to be devised). The following are the requirement to be met:

1. The growing connectivity issue of the proliferating devices or appliances which tap into the web.

2. Security and privacy issues.

3. Seamless integrating of new technologies (wired and wireless) which have evolved after the advent of internet and

4. A socially responsible network which could cater to emergencies and even deliver location specific content.

1.1 Literature

The literature related to the Topology, Unique Addressing problem for each and every node in a network, Graph labellings, Graph Theory Approach for Unique Address Assignment Problem, Dominating sets and Metric Dimension in Graphs, Clustering Network, Land Marks in the Graph, Routing In Interconnection Networks are addressed.

1.1.1 Topology

A large number of topologies [1] have been proposed and studied for multicomputer interconnection networks. The Existing static interconnection networks include linear arrays, rings, meshes, complete binary trees, X-trees, full-ringed binary trees, tree machines, pyramids, fat trees, hyper cubes [2], meshes of trees, cube-connected cycles [3], cross product networks [4,5,6] and de Bruijn networks. A 3 dimension hypercube consisting of 8 nodes has a diameter 3 where
as a Peterson graph having a degree 3 and 10 nodes has a diameter 2. In spite of this the larger diameter hypercube is preferred to Petersen because of its other advantages such as symmetry, regularity, high fault-tolerance, logarithmic degree and diameter, self routing and simple broadcasting [7] schemes. However this offers scope for a look at a novel interconnection [8,9,10,11].

1.1.2 Uniquely Addressing Problem

For a large network of Transmitters/Receivers (Nodes/Vertex) spread out in a planar region, the addressing problem is to assign a numerical value, representing a hop count (Tuple), to each transmitter. The addressing assigned to nearby transmitters should overcome some separation constraints so as to avoid interference. The goal is to uniquely address each and every node in a network, so that it is possible to form a non-overlapping cluster which helps routing algorithm for an efficient path planning.

1.1.3 Graph Labeling

Most graph labeling methods trace their origin to one introduced by Rosa [12], [31] in 1967, or one given by Graham and Sloane [13]. Rosa [12] called a function $f$ a $\beta$-valuation of a graph $G$ with $q$ edges if $f$ is an injection from the vertices of $G$ to the set $\{0, 1, 2, \ldots, q\}$ such that, when each edge $xy$ is assigned the label $|f(x) - f(y)|$, the resulting edge labels are distinct. Golomb [14] subsequently called such labellings graceful and this is now the popular term. Although an unpublished result of Erdos says that most graphs are not graceful, most graphs that have some sort of regularity of structure are graceful.

Labeled graphs serve as useful models for a broad range of applications such as: coding theory, X-ray crystallography, radar, astronomy, circuit design, communication network addressing and data base management.
1.1.4 Graph Theory Approach for Unique Address Assignment Problem

The metric dimension problem, introduced by F. Harary in 1976 [15], has been recently widely investigated [14],[16],[17],[18],[19]. It arises in many diverse areas including the network discovery and verification, the robot navigation, connected joints in graphs, chemistry, etc.

Given a simple connected undirected graph $G = (V,E)$, for $u,v \in V, d(u,v)$ denotes the distance between $u$ and $v$ in $G$, i.e. the length of a shortest $u-v$ path. A vertex $x$ of the graph $G$ is said to resolve two vertices $u$ and $v$ of $G$ if $d(x,u) \neq d(x,v)$. An ordered vertex set $S = \{x_1, x_2, \ldots, x_k\}$ of $G$ is a resolving set of $G$ if every two distinct vertices of $G$ are resolved by some vertex of $S$. Given a vertex $t$, the $k$-tuple $r(t,S) = (d(t,x_1), d(t,x_2), \ldots, d(t,x_k))$ is called the vector of metric coordinates of $t$ with respect to $S$. A metric basis of $G$ is a resolving set of the minimum cardinality.

The metric dimension of $G$, denoted by $\beta(G)$, is the cardinality of its metric basis.

Each network is a graph in which vertices are transmitters and any pair of vertices that may interfere is represented by an edge in the graph.

Graph coloring is one such assignment. In proper coloring we assign the colors to the vertices in such a way that the adjacent vertices receive the distinct colors. In other words, by representing the colors by positive integers, we have a proper coloring in function $f : V(G) \rightarrow \mathbb{Z}^+$, such that $|f(u) - f(v)| \geq 1$, whenever $d(u,v) = 1$.

Griggs [19] proposed the graph-theoretic analogue of the distance-constrained channel assignment problem, and he generalized it to permit $p$ levels of interference. Specifically, given integers $k_1, k_2, \ldots, k_p \geq 0$, called separations, we have a $L(k_1, k_2, \ldots, k_p)$ labeling of a graph $G$ is an assignment of non-negative integers $f(v)$ to the vertices $v$ of $G$, such that $|f(u) - f(v)| \geq k_i$, if $u$ and $v$ are at dis-
tance \( i \) in \( G \), we say that labeling \( f \) belongs to the set \( L(k_1, k_2, k_3, \ldots, k_p)(G) \).

We denote by \( \lambda(G : k_1, k_2, \ldots, k_p) \) the minimum span over such \( f \), where the span is the difference between the largest and the smallest labels \( f(v) \). Many authors have subsequently contributed to the literature on these labellings. The entire literature is available in the recent article by Jerald R Griggs and Xiaohua Teresa Jin [20].

1.1.5 Dominating sets and Metric Dimension in Graphs

A subset \( D \) of \( V \) such that every vertex in \( V - D \) is adjacent to at least one vertex in \( D \), is called a dominating set and the minimum cardinality of a dominating set is called domination number of the graph \( G \) and is denoted by \( \gamma(G) \).

For safeguard analysis of a facility, such as a fire protection, study of nuclear power plants, the facility can be modeled by a graph or network. For such applications, a vertex can represent a room, hallway, stairwell, courtyard, and etc. Each edge can connect two areas that are either physically adjacent or within sight or hearing each other. One primary function of a safeguard system is 'direction' of some object—perhaps detection of a fire, or detection of an intruder, such as a saboteur. Suppose we have a detection device located at a vertex and the device can supply three outputs,

1. There is an object at that vertex.
2. There is an object at one of the vertices adjacent to that vertex, or
3. There is no object at that vertex or any adjacent vertex.

It is necessary to determine a collection of vertices at which the detection devise can be placed so that if there is an object at any vertex in the graph, it can be detected, and its position uniquely identified. In order to detect an object, which might be at any vertex in \( V(G) \), it is necessary to have a dominating
set. The additional problem of uniquely identifying the location of the object requires a locating feature.

This concept served as motivation for the investigation of the new graph invariant called metric dimension (locating number). For any graph $G = (V, E)$, the metric basis (locating set) of $G$ is the subset $S$ of $V$ having the property that for each pair of distinct vertices $u, v$ in $V$ there exists a $w$ in $S$ such that $d(u, w) \neq d(v, w)$ in $G$. A minimal cardinality of a metric basis (locating set) is called metric dimension (or locating number) of $G$. The concept of metric dimension is introduced by F. Harary and R. A. Melter [15] and as a locating number independently by P. J. Slater [21] in 1975.

P. J. Slater [22], called a dominating set $D$ of a graph $G$ as a locating dominating set of $G$ if for each pair of vertices $u, v$ in $V - D$, $N_D(u) \neq N_D(v)$, where $N_D(u) = N(u) \cap D$. The minimum cardinality of a locating dominating set is called the locating domination number $\gamma_L(G)$. Metric dimensions and locating dominating sets of certain classes of graphs were studied in [15], [16], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35].

1.1.6 Clustering Network

Clustering approach is used ideally to distribute the traffic load evenly across the network, for maximizing its potential [36]. However disparities in network topologies, routing algorithm and traffic pattern make it difficult to achieve a perfect balance. As a result, there may be some links that are more heavily loaded and potentially more susceptible to congestion. On the other hand there may be group of nodes which are connected by links that are less heavily loaded and have higher aggregate bandwidth. Such group of nodes are referred to affinity clusters. These could be potentially exploited to give better performance and also provide us insight into the structure of the network. Partitioning a network into clusters of richly connected nodes serves two important purposes:
1. More heavily loaded inter-cluster links can be identified (potential bottleneck links and the points in the network where additional bandwidth is likely to improve the performance)

2. Simple and meaningful characterization Matrices can be measured on the clusters.

The challenge here is not only how we can find the cluster under decentralized management, but also how accurate and efficient the algorithms are that are used for finding good clusters, how accurate such distributed node clustering schemes can be, when compared with centralized clustering algorithm, how accurate and beneficial these mechanisms are when compared with the option of re-clustering of the network on the entrance and exit of each node. We believe that every such network exhibits some unique structural properties and discovering these network structures is crucial to efficient data discovery, node lookup and communications. There has been considerable research in the algorithm community towards addressing the problem of clustering the node in direct and undirected graphs. Although some researches have applied clustering information to address certain key problems in decentralized P2P network [37], very few of them have studied the problem of discovering and maintaining the node clusters in P2P system.

Based on some properties of good clustering, a vast number of clustering algorithms have been proposed. The two most popular schemes are the K-path clustering algorithm and the MCL. Like most existing graph clustering algorithms, both of them assume that the global information about the entire graph (that is the number of vertices and number of edges and their connectivity) is available in one central location. Unfortunately, none of the networks maintain their complete up-to-date connectivity information. Therefore, the need is to design schemes that can cluster the nodes of a network in a completely distributed and decentralized manner assuming that each node has a limited view
of the entire network. The connectivity-based node clustering technique is one such interesting and important network structure that can be utilized in various ways to improve the quality of service of application running on these networks (P2P). A connectivity based node clustering can be defined as a partition of network nodes into one or more groups based on their connectivity.

All clustering algorithms proposed to date so far, have some or other weaknesses. Some methods rely on node IDs in eliminating potential redundant broadcasting nodes or in defining priorities. Those approaches suffer from the fact that they can not detect all possible eliminations, because ordering based on node ID (or node weight) prevents this. As a consequence they incur significant excessive retransmissions. Some methods do not fully exploit the compiled information. For instance, the use of the node degree as its priority when deciding whether it will become a cluster head might not result in the best local decision. Finally, some methods create a lot of clusters, or require excessive communication cost. The basic novel clustering algorithm for wired networks complies with all requirements described below.

- It is localized, thus distributed, it can exploit 1-cluster, 2-cluster or k-cluster different trade off in efficiency versus network diameter and control overhead
- The cluster heads are estimated on the basis of the originator node, which maintains an unique ID
- Allows fast network clustering, generating the unique ID
- Identification of reference point (landmarks, workstations, cluster heads, servers, central processors) and address generation while routing to create hierarchical level

Network and data management are completely decentralized. Individual nodes have limited knowledge about the structure of the network [38], [39], [40], [41].
There is a need for better algorithm for clustering.

### 1.1.7 Land Marks in the Graph

In a given network (graph) for a node to determine its location uniquely, it requires a minimum number of landmarks and their well defined position. The set of nodes where the land marks are placed is called the metric basis of the graph and the number of landmarks required to identify uniquely is called the metric dimension of the graph \[12, 13, 30\]. The problem of finding the metric dimensions of a graph was first reported by Harary and Melter\[15\]. They gave the characterization, for the metric dimension of the trees. However their proof has an error or more specifically the proof of lemma 1 has an error. The formal definition of the notion of metric dimension is as follows. The metric dimension of graph \( G = (V, E) \) is the cardinality of the smallest subset \( s \subseteq V \), such that, for each pair of vertices \( u, v \in V \), there is a \( w \in s \) such that the length of the shortest path from \( w \) to \( u \) is different from the length of the shortest path from \( w \) to \( v \).

The ”coordinates ” is associated with each node based on the distances (Hop Counts) from the node to the landmarks, with the aim of picking enough landmarks, so that each node has a unique tuple of coordinates. Let \( G = (V, E) \) be a connected undirected graph. A ”coordinate system” on \( G \) is defined as follows. A set of nodes are picked up as the metric basis and each node in the basis corresponds to a landmark. For each landmark, the coordinate of each node \( v \in V \) in the corresponding ”dimension” is equal to the length of a shortest path from the landmark to \( v \). Thus for a metric basis, each node has a vector of coordinates, a tuple of non negative integers, specifying the distances to that node from the nodes in the basis. A metric basis for connected graph \( G \) is the smallest set \( s \) of points \( G \), such that no two points have the same distances to all points of \( s \). The metric dimension of \( G \) is then the cardinality \( s \).
1.1.8 Routing In Interconnection Networks

In the case of irregular networks, it is relatively difficult to develop an efficient algorithm. Such a routing algorithm should provide deadlock freedom, low network latency, and high link utilization. Furthermore, the routing algorithm should be efficiently implemented in a communication switch. Some routing algorithms for irregular networks have been developed [44]. They can avoid deadlock, but these methods rely on a large hardware routing table in the communication switch.

As the network grows, the size of routing table grows proportionally. Hence, the complexity of implementing the switch increases. Besides, packet traversing the switch suffers from long latency of making routing decision. High-throughput low-latency interconnection networks are suitable for workstation clusters. An interconnection network can be constructed in different topologies. Typically, interconnection networks with regular topologies, such as mesh, torus, and hypercube, are less scalable for workstation clusters than those with irregular topologies. The TRAIN routing algorithm is deadlock-free and requires no routing tables in the switch. This routing algorithm is based on a spanning tree which is constructed in the network. A spanning tree is a subset of the network. However there are many ways of constructing a spanning tree, and each construction yields spanning tree with varying path length, degree and cost. Thus there is a need for a better routing algorithm.

1.2 Background and Motivation

1.2.1 Clustering based on Identity Number

In ad hoc networks, nodes are distributed randomly and they are identified by their unique IDs. In such a network, the execution of a leader election algorithm would result in the identification of only one node. If that node is the only node
monitoring all other nodes, then the system no longer remains distributed but becomes centralized. Such a process will result in a single cluster and would require tremendous amount of transmission and computation power. Hence, we have to use a modified version of the leader election scheme. Instead of choosing only one leader, multiple nodes can be elected to perform the duties of leader. Each of the leaders would be a leader with respect to a subset of the nodes—that is, it would have dominance over a set of nodes. Each leader would know who belongs to the set. This gives rise to the concept of clustering.

A cluster is similar to a "group" in distributed systems. The nodes within the cluster are considered to be the members of a cluster head, a node has to lie within the transmission range of that cluster head. However, it is also possible that even if a node is within the transmission range of a cluster head, it may not belong to that cluster head. On the other hand, if a node falls within the transmission range of more than one cluster head, it is called a gateway node. A distributed gateway is a pair of neighboring nodes from different clusters physically located nearest to each other even though their clusters do not overlap. Most clusterhead election algorithms require that each node is a member of one cluster head and no two cluster heads are adjacent to each other. Thus, two cluster heads cannot be one-hop neighbors. Therefore, ad hoc networks make use of multiple clusters and require multiple cluster heads.

1.2.2 Routing

Routing is one of the most important reasons behind clustering of the network. Routing essentially means finding a path to send a packet from a source node to a destination node via single or multiple hops. Mobile ad hoc networks present unique challenges to routing protocols due to the dynamic nature of the network. If a node moves out of the transmission range of all its neighbors, it
becomes unreachable. In addition, the network may become partitioned into clusters where the inter-cluster communication is not possible. Many fault-tolerant protocols can recover from being disconnected by treating the node as a failed station, but they may have trouble handling a partitioned network, especially if the nodes cannot reach a consensus on which partition should be considered “disconnected”.

A more pervasive problem is the lack of stability in the network topology. The optimal route between two nodes is constantly changing as the end points affiliate between clusters and the dominant set changes. If the lifetime of a route can be measured accurately and efficiently the routes may last till the end of the transmission of packets. By maintaining the route lifetimes, we ultimately achieve to keep the network in stable condition as long as possible. However, if the lifetime of a route between two nodes is shorter than the life time required to transmit packets through the route due to node and/or link failure, the packets may be dropped in the middle. Therefore, each routing protocol provides two main mechanisms, namely route discovery and route recovery/maintenance. As the name suggests, the route discovery finds possible routes between any given source and the destination pair, based on certain quality of service criteria. Route recovery provides various ways of ensuring the successful transmission of packets after route breakage. The routes need to be recovered as early and efficiently as possible so that the packet delays can be minimized.

Routing in multi-hop packet radio networks was based on the shortest path algorithms [45]. The Distributed Bellman-Ford algorithm [46] is one such algorithm currently in use in routing algorithm [47]. In general, the routing protocols can be divided into two main categories in relation to the route discovery and maintenance mechanism that they provide, namely proactive and reactive. Proactive, also known as table-driven, protocols maintain up-to-date routing tables for each node in the network. Since every node knows how to reach the rest of the nodes, the route discovery mechanism is very simplified. However,
the route recovery and maintenance phase becomes much more involved due to propagating the updates to every node in the system. If the nodes in the network are highly mobile, the routing table updates become computationally expensive. Different nodes may still have inconsistent routing tables due to incomplete propagation of dominant set changes. Thus, a packet may temporarily enter a routing loop until update is correctly processed. The destination-Sequenced distance -Vector (DSDV) \[18\] is the most commonly used proactive routing protocol. The clustered gateway switching routing (CGSR) \[49\], which uses DSDV as the underlying routing scheme, and the Wireless Routing Protocol (WRP) \[50\] are other problems.

Reactive, also known as source initiated, protocols discover the routes as the source node initiates the route discovery process. The nodes do not maintain routing table.

1.3 Problem Statement

The designing of an efficient routing algorithm for a given network is in focus. Routes between two hosts in the network may consist of hops through other hosts in the network. The task of finding and maintaining routes in a network is difficult if the network is centralized, because of the maintenance of the routing table at each and every node. The flooding scheme is the most reliable for sending data packets. However, since the link channel and battery power resources are very scarce, more efficient schemes must be devised. These schemes require up to date information about the location of nodes. Storage is not a critical issue since memory continues to get less expensive each year. The savings in communication bandwidth and energy come from reporting only to nodes that need a particular information. To reduce the transmission overhead for the update of routing tables after topological changes, it was proposed to divide all nodes into clusters. The overhead of cluster formation and maintenance
can not be ignored. In the general cluster-based schemes for ad hoc networks, clusters are formed at first, and one cluster head (CH) is elected for each cluster, in the fully distributed fashion. In cluster-based approaches, the sender must know the location information of the cluster within which the destination is located.

The routing algorithm may consist of routing from source to its CH, from the CH to the CH of destination node, and from the latter node to the destination. Communication between CHs involves intermediate nodes in their clusters. To reduce the power consumption in CH nodes, the information about all CHs may be replicated in all the nodes of the network. Therefore each node stores the information about all the clusters (more precisely, about CHs) in the network. Each node knows the content (i.e., the list of nodes) only for its own cluster. The sender may forward directly toward s destinations CH, and does not need to consult its CH. Moreover, the routing paths must necessarily have to pass through any of the CHs, since the message can be rerouted toward the next cluster as soon as it enters any of the clusters.

A hierarchical network structure is an effective way to organize a network comprising a large number of nodes. In a single hierarchy, nodes are divided into clusters, which may or may not have cluster heads. It is suitable for networks with a few hundred nodes. A multi-level hierarchy has nodes organized in a tree-like fashion with several levels of cluster heads. A three level hierarchy employs ordinary nodes, cluster heads and super-cluster heads, and is suitable for networks with a few thousand nodes.

With the objective of developing a Routing algorithm based on the location information generated for each and every node, the following list of research objectives is included as the scope of the thesis: The intermediary objectives to be addressed are as follows

1. Generate unique location information for each and every node in a network
2. Characterize the graphs with metric dimension two

3. Introduce the definition of a cluster

4. Characterize the graphs with cluster dimension two

5. Develop a clustered network (non-overlapping clusters)

6. Develop a routing algorithm for a clustered network (LMUAC)

1.4 Organization of the Thesis

This thesis addresses some fundamental issues such as uniquely addressing problems, clustering, routing and path planning in networks. The techniques and algorithms which are used to construct clusters and to maintain the unique clustering Hop ID for the nodes in a network so that communication between the source and destination is supported without disruption. The rest of the thesis is organised as follows.

Chapter 2 deals with introduction of Graph Theory approach for communication and introduction to metric dimension two (Unique addressing) results.

Chapter 3 deals with characterization of metric dimension two. The distance partition of vertex set of a graph $G$ with reference to a vertex in it is defined and with the help of the same, a graph with metric dimension two (i.e. $\beta(G) = 2$) is characterized (The result of this chapter is published in Journal of AADM 2009 and a part is published in the proceedings of ICAMCS 2009).

In chapter 4, the Cluster Dimension of a network is defined as the minimum cardinality of a subset $S$ of the set of nodes having the property that for any two distinct nodes $x$ and $y$, there exist the node $s_1$, $s_2$ (need not be distinct) in $S$ so that $|d(x, s_1) - d(y, s_1)| \geq 1$ and $d(x, s_2) < d(x, s)$ for all $s \in S - \{s_2\}$ (The result is communicated to journal IJAMCS and a part of this work is published in the proceedings of MWOCN 2009). In this chapter, the sufficient conditions for a graph of cluster dimension $n$ and a tight upper bound for the number of nodes of a network with prescribed dimension in terms of diameter are obtained.
In chapter 5, the distance partition of vertex set of a graph $G$ with reference to a vertex in it is defined and with the help of the same, a graph with cluster dimension two (i.e. $\beta_c(G) = 2$) is characterized. In the process, a polynomial time algorithm is developed which verifies if the cluster dimension of a given graph $G$ is two. The same algorithm explores all cluster bases of graph $G$ whenever $\beta_c(G) = 2$ (The result is published in the proceedings of MWOCN 2009 and a part of the result is communicated to GJPAM).

In Chapter 6, a Land Marks for Unique Addressing(LMUA) algorithm is developed to generate unique ID for each and every node which leads to the formation of overlapping/Non-overlapping clusters based on unique ID. The LMUA algorithm constructs overlapping clusters which is a drawback, To strictly construct the no-overlapping clusters the concept of clustering is introduced based on unique hop ID. Based on the clustering concept a Land Mark for Unique Addressing and Clustering Algorithm(LMUAC) is developed. With the development of LMUAC Algorithm the non-overlapping clusters are constructed, with the help of the same the classification of those nodes into cluster heads, Member Nodes, Gate way nodes is carried out. The algorithm is developed to generate the Hierarchical code for the cluster heads to operate in the level one hierarchy for wireless communication switching. Whether the expansion of the existing network can be performed or not with out modifying the cost of adding the cluster head is presented.

In chapter 7, the design of LMUAC Routing scheme for an efficient path planning from source to destination by using the Unique clustering Hop ID is developed. Using LUMAC algorithm, the upper Bound expression in terms of path length, Maximum Link Utilization and Memory space requirement are calculated and show that all the above perform better in the clustered network(The results of this chapter is communicated for MWOCN2010 is accepted for oral presentation). Finally, chapter 8 concludes this thesis with directions to future research.