สำนักงานวิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER I

INTRODUCTION

1.1 Introduction and Problem Review

The shortest path problem is one of the most important generic problems in such fields as operations research (OR), computer science (CS), and computer engineering (CE). One of the reasons for this is that any combinatorial optimization problems can be formulated as a shortest path problem (SPP) . Thus, this class of problems in these days is extremely large and includes numerous practical situations. For examples, SPP is applied in the inventory system, logistics and transportation system, computer networking system, etc. However, these systems are altered from the legacy environment. Therefore, the current criterion cannot be solved by the existing shortest path method.

Since the mobile environment has been influencing to human life style, therefore, various researches have considered the shortest-path and nearest neighbor queries in the geo-positioning e.g., road network. Unfortunately, the existing solutions have not considered the situation of the network congestion; a primary solution that had been applied to the SPP is Euclidean Space while the nearest neighbor queries in spatial networks have been proposed by [1-6]. On the other hand, [7-10] considers the static travel-time in the road network and geo-position locator concerning on path selection of spatial network query. However, these existing techniques considered the weighted cost of paths using static information of path distance, or immediate travel-time. So, it is not enough to perform a future traveling plan when the travelling path is congesting.

The important keys to satisfy the needs of SPP applications that require fast response are relied on two levels. Firstly, the collecting data must be the most up-to-date weighted data from data sources which are usually mobile. Secondly, the collected data must be organized and calculated by a suitable algorithm. Nevertheless, all of these processes have to accomplish in an acceptable response time. Therefore, the first process where the up-to-date data is required from every source must be transferred from mobile stations with a high throughput and low delay before the second process starts.

In order to collect data from mobile stations with a high throughput and low delay in the wireless environment, the hop-to-hop data transfer mechanism is concerned. This research focuses on the wireless network that capable for important requirements of the most SPP applications i.e. metropolitan coverage, wide bandwidth support for high throughput data transfer, acceptable end-to-end delay, and practical to implement for the real world business of network providers. One additional characteristic of the wireless network is that it must be a low cost coverage extendable by multi-hop relay without the throughput and delay compromised.

Multi-hop wireless access network, IEEE 802.16j [11-13], is the subject of much interest at present. In the last few years, this network has moved from the domain of a research interest to a significant impact on the commercial world. There are two key indicators of this change which are the increment of standardization activities relating to multi-hop networking, and the growth of companies with product portfolios based on such technology.

After considering the benefit of IEEE 802.16j technology mentioned above, transferring objects, such as data packet, over networks can be counted as a SPP then the first success factor described previously must be obtained. Thus, one part of this research will focus on the IEEE 802.16j multi-hop relay network to solve the data transfer problem.

Since data transferring mechanism in IEEE 802.16j multi-hop relay network must deal with the path finding problem, thus, the second success factor of SPP will be considered in this research as same as the first success factor mentioned above. In order to response to this problem, after data are collected properly, it will be organized, updated, and performed path calculation with a new and fast algorithm. Finally, the most suitable transferring path will be chosen and presented. In order to perform such processes, the system needs several components i.e. a datacenter function, a mobile unit function and communication protocol which have to work cooperate with an improved data structure and adaptive path finding algorithm that is highly needed.

1.1.1 Problem 1: Data transfer problem

This subsection describes how data are collected from mobile stations with high throughput and delay sensitive requirements, over the mobile distributed architecture. For above requirements, IEEE 802.182j seems to be the most suitable with the requirements.

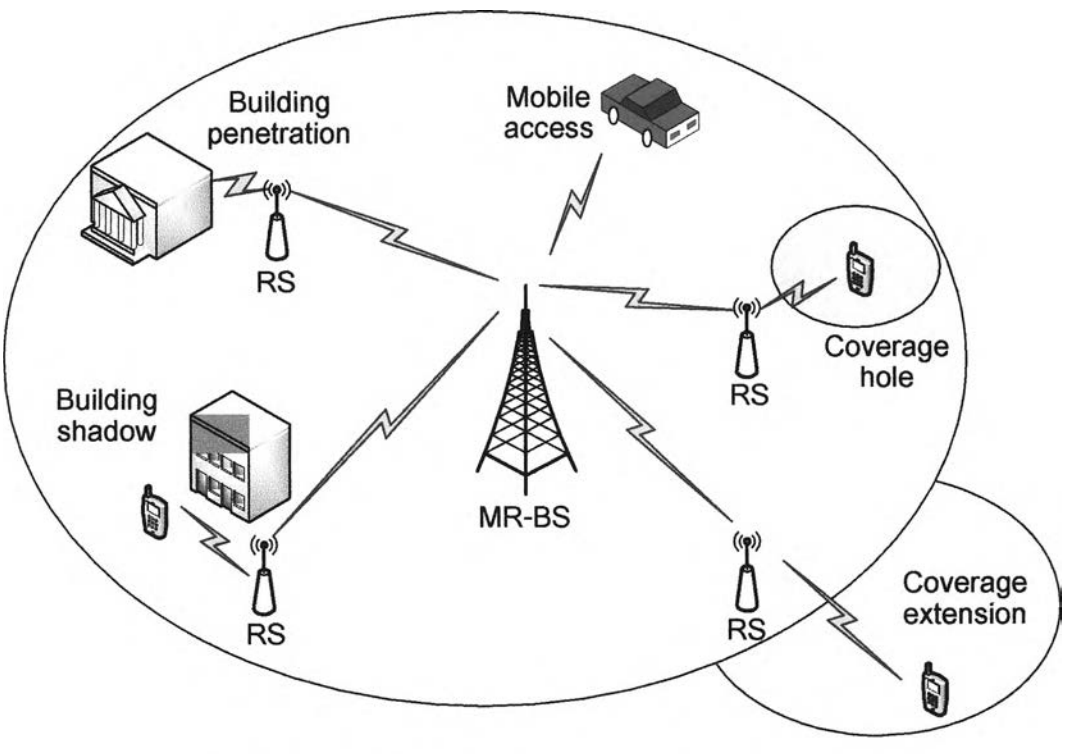

Figure 1.1: IEEE 802.16j multi-hop network applications.

The WiMAX technology is becoming more popular [14-21] as the sharply increasing numbers of service providers deploying WiMAX for their new wireless broadband services. The newly formed IEEE 802.16j [11-12, 18-22] working group is to focus on mobile multi-hop relay networks that enable multi-hop communication in IEEE 802.16e networks [13, 18]. In multi-hop relay network, one or more mobile stations (MS), or subscriber stations (SS) may communicate with relay stations (RS) instead of communicating directly with multi-hop relay support base stations (MR-BS). IEEE 802.16j is an amendment to the IEEE 802.16 broadband wireless access standard to enable the operation of multi-hop relay stations. It aims to enhance the coverage, per user throughput and system capacity of IEEE 802.16e. Comparing with a base station (BS), the RS does not need a wire-line backhaul; additionally, it has much lower hardware complexity. Thus, using RSs can significantly reduce the deployment cost of the system [14, 15, 16, 21, 22].

Unfortunately, In the IEEE 802.16j Multi-hop Relay Network, those MSs connect to the RS located at two or more hops away from the MR-BS, are suffered from throughput degradation and end-to-end delay increase [23-44]. As the number of RS hops increase, the degradation and the delay grow. Therefore, these problems are two major problems in the IEEE 802.16j Multi-hop Relay Network [24-28, 31, 38, 40].

There are two main reasons that are the cause of highly degradation of both throughput and delay in the IEEE 802.16j network. First is the multi-hop characteristic that causes an increasing delay and degrading the throughput when the number of RS hops increases. And second, the signal interfere limitation for each RS hop, causes a lower utilization in the relay zone when the number of RS hops increases.

The frame structure standard of IEEE 802.16j [11-12] supports two approaches. The first approach allows one or more RS or MR-BS frames to be grouped into a multi-frame with a repeating pattern of allocated relay zones. This approach causes limitation to the relay zone utilization, especially on over the 3 hops scenarios. The second approach enables a single-frame structure consisting of more than one relay zones. This approach is appropriated to accommodate the network coding. However, the more RS hops increase, the more relay zones of RSs and MR-BS need to stay in the idle mode. This will lead to lower the frame structure utilization and cause throughput degradation.

หอสมุดกลาง สำนักงานวิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

It is the truth that the SSs who connect to the RS with two or more hops located in a long distance measured from the MR-BS, are suffered from the bottle-neck of multi-hop, throughput degradation and increasing of end-to-end delay. These interesting problems may cause some serious situation for particular applications over the wireless multi-hop network. Therefore, numerous solutions have been proposed [6, 21-31, 34-35, 38, 41-46] but the problems are partially solved and remain challenge.

1.1.2 Problem 2: Path finding problem : how to response to the shortest path query fast and accuracy

New classes of genuine SPP are becoming very important these days in connection with practical applications [1, 3-10], such as on-line computing of driving directions, road traffic management, logistic planning, transportation planning, network routing and telecommunication planning. These problems need a large adaptive complex weighted-graph to represent the entire processes. However, using the existing method to solve these weighted-graph problems with a large number of nodes will be very time-consuming to obtain a proper answer. Besides, some presented outputs may not be a real optimal solution.

The one classical problem was selecting the best path to travel from one point to another by least travel-time as possible. This is the significant interest under the problem of individual travel and business logistic planning [1, 7-8, 10, 47-95]. In order to solve this problem, the required preliminary information was a digital map of possible transportation path (e.g. road map for land transportation, marine/watercourse map for waterway transportation). Then, calculations have been performed to find the shortest-path from the source to the destination according to the digital map, as the representation of the directed weight graph.

The digital map and GPS are widely used in combination with the wireless communication system. The GPS with the digital map enable many navigation services extended from a mobile unit locator service. Thus, high accurate information for navigation services can be obtained from mobile units by the datacenter.

In the real situation, especially in a high traffic congestion environment, various factors can influence to the travel-time to the destination (e.g., congestion, driving behavior, vehicle limitation and special congestion event). Unfortunately, the existing methods [1, 7-8, 10, 47-88] do not cover this problem.

Thus, in the real situation, the congestion can make a large impact on the shortest-path selection in term of the travel-time. Beyond that, period of time during a day and the days during a week can make a big difference on congestion of each road. For instance, in weekday rush hours, the road approach downtown will be very busy. On another hand, the same road will almost empty by late of the night time.

1.1.3 Problem domain

The focusing domain in this research is in the area of traffic management where SPP is the primary concerned. Thus, some main components related to the traffic management are the transferred data over networks, the datacenter, mobile unit (MSs), BSs, and RSs. Therefore, this thesis will consider all necessary criteria that support performance of traffic management process. Thus, the proposed solutions in this thesis include;

- Multi-hop wireless network

  The considered wireless network in this thesis is the network coding and a frame structure design on the IEEE 802.16j Mobile Multi-hop Network.

- Data structure

  This thesis considers the data structure format and organization on the datacenter.

- Algorithm

  The SPP algorithm in this thesis focuses on the datacenter process based on dynamic congestion situation on the traveling paths.

1.2 State of problem

1.2.1 Data transfer problem

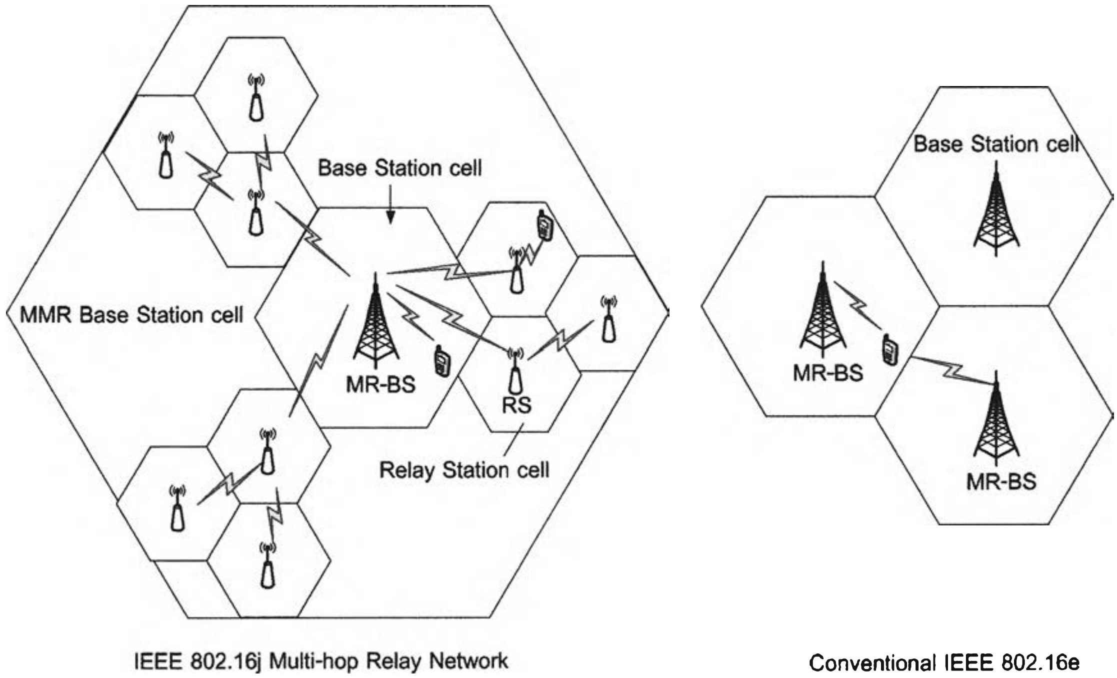

Figure 1.2: IEEE 802.16j Multi-hop Relay Network and conventional IEEE 802.16e.

The IEEE 802.16j multi-hop relay network is based on the scheduling scheme. In centralize scheduling mode, all SSs and RS are determined by DL/UL map from the scheduler at a MR-BS, when they transmit, receive or idle. However, there is a slightly difference in the distributed scheduling mode, an RS can create their own DL/UL map and distribute to their connected SSs. The two importance problems in the IEEE 802.16j multi-hop relay network is that every SS which connects to the RS with 2 or more hops away from the MR-BS will be suffered from the bottle-neck of multi-hop. As a consequence, end-to-end throughput degradation and increasing of end-to-end delay occur. For the throughput degradation problem, it is effect from the saturated network situation in the relay zone. The relay zone is a channel that allows the communication between the MS-BS and the RS or the RS to the subordinate-RS.

The same problem with other wireless multi-hop networks also occurred in the 802.16j multi-hop relay network, due to the multi-hop network aligns like the line topology. Since each relay has obligation to forward traffic to both UL and DL, this situation creates a bottle-neck in the relay link on the multi-hop network. in a high traffic loads, the relay link is the cause of saturation of all multi-hop networks. So, the throughput will largely degrade [26-46], and even more on the farer hops relay.

Considering all SSs that connected to the RS that two or more hops away from the MR-BS. These SSs will be suffered from the bottle-neck of multi-hop end-to-end throughput degradation. These highly degradation of both throughput and delay in IEEE 802.16j network are the effects of two factors. The first factor is the multi-hop characteristic because it causes an increasing delay and degrading the throughput when each number of RS hops was increased. The second factor is the limitation of signal interference of each RS hop. This limitation leads to a lower utilization in the relay zone when the number of RS hops increased.

1.2.1.1 Multi-hop throughput degradation

The number of bits passing through a communication channel in a certain period of time, called as throughput, is very important because some services require a high throughput rate in order to prevent a critical loss. Therefore, the factors that affect to the number of throughput rate must be seriously considered. In IEEE 802.i6j multi-hop relay network, there are two important factors: Multi-hop characteristic and Signal interference limitation, as described below.

1.2.1.1.1 Throughput degradation factors: Multi-hop characteristic

The saturated network in other wireless multi-hop network had been studies in the past [23-44]. When the offered load increase through the multi-hop wireless network, the end-to-end throughput will respectively increase until the peak throughput point reaches. The peak throughput point is named as the optimal offered load, and the offered load beyond this point is unsustainable; moreover, the throughput will start dropping. In the IEEE 802.16j multi-hop relay network, this optimal offered load also exists. In the general case, this optimal offered load point is also controlled by the bottle-neck situation of the relay zone too.

Start with bottle-neck situation that will occurred to the relay zone between a RS and a MR-BS/super-ordinate RS, when the traffic load of all SSs that connected to the considered RS and their sub-ordinate RSs exceed the maximum traffic loads over the relay zone. In a general case in the line chain of RSs, the bottle-neck situation is usually occurred at the relay link between a MR-BS and a RS1. This bottle-neck at the relay zone becomes another limitation to the offered load. The end-to-end throughput in the IEEE 802.16j multi-hop relay network is depended on the minimum load between the optimal offered load point and bottle-neck point. Thus, by applying the proposed NC-BR, the network coding can support RSs in transmitting higher data rate over the relay link. Consequently, both optimal offered load point and the bottle-neck point are relieved. More details are described in the next section.

1.2.1.1.2 Throughput degradation factors: Signal interference limitation

On multi-hop relay network, RSs are connected in a chain topology alike. In a simple configuration, each RS transmitting range is just about to cover its super-ordinate, sub-ordinate RS and connected MS. As the result that all RSs use wireless as an backhaul when each RS is in the transmitting mode, their super-ordinate and sub-ordinate RS have to be in the receiving mode to avoid the signal interference. Therefore, the MR-BS scheduler has an obligation to determine transmission that turns for each RS. Accordingly, when the number of hops increases, the transmission time interval for each RS will be arisen, and finally, the network utilization will be dropped. In order to unsolved this problem, a study of reusing frequency channels had been proposed [45]. Unluckily, the result indicates a small improvement of utilization.

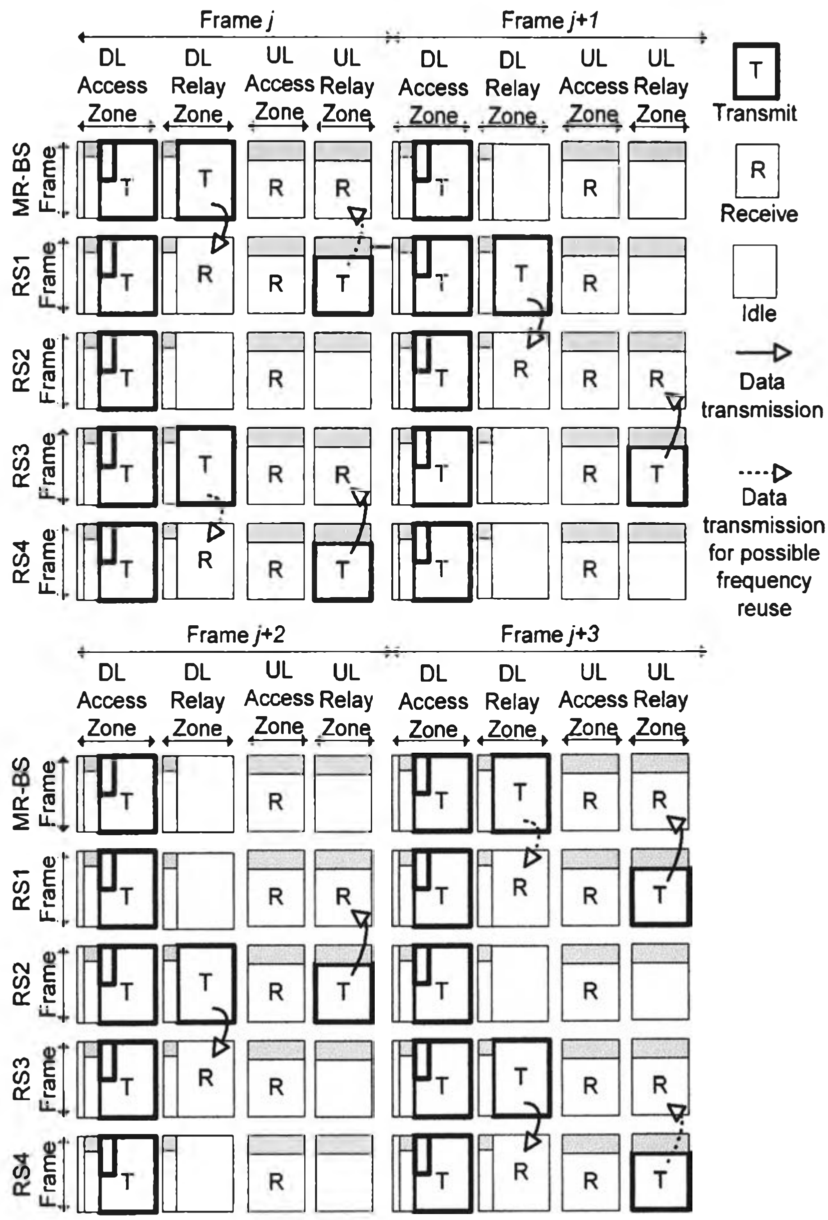

Figure 1.3: Transmission sequence in non-transparent frame structure of 5-hops IEEE 802.16j multi-hop network.

Figure 1.3 shows a simple non-transparent frame structure transmission sequence for the IEEE 802.16j multi-hop network under 5 hops, according to the latest standard document of IEEE 802.16j [11-13].

1.2.1.2 Multi-hop delay increases

Beside the throughput degradation problem, another issue that must be focused is the end-to-end delay. In the 802.16j multi-hop network, similar to various wireless multi-hop networks, each hop of the relay is the cause of an additional time for the end-to-end delay when a traffic flow through [23-44, 46]. Besides, the minimum of the additional delay is the size of frame duration of a RS. So, the connected SSs are also suffered from this additional delay, according to the fact that many hops away from the BS, many delays are added.

Thus, to make a clear understanding of the end-to-end delay over the IEEE 802.16j multi-hop network, two main delay values must be counted: a multi-hop delay, and a traffic flow delay. The details of these factors are described as follows.

1.2.1.2.1 Delay increases factor: Multi-hop delay

A multi-hop delay, the traffic flow transmitted thru the number of the MR-BS and RSs until reaching the MS, is related to the wireless multi-hop characteristic. At each RS, the time that the RS needs to process a set of data and forward to the destination is called an operational delay. While the time for a signal transferring between RSs and from MR-BS/RSs to their connected MS are called a propagation delay on a relay link and a propagation delay on an access link, respectively.

1.2.1.2.2 Delay increases factor: Traffic flow delay

When a network is saturated, RSs are unable to clear all data in their forward queues. In this case, the data becomes a backlog in the RS forward queue, and the end-to-end delay is certainly increased. This type of delay stated as a traffic flow delay. This traffic flow delay is occurred in a relay link and associated to the bottle-neck position. In the relay link, multiple traffic flows of all sub-ordinate RSs are sharing the same relay link bandwidth which is so inadequate. The traffic flow delay can be defined by the time that needed to be cleared out from the backlog data thru a specific relay link at each RS.

1.2.2 Path finding problem: Shortest-path problem

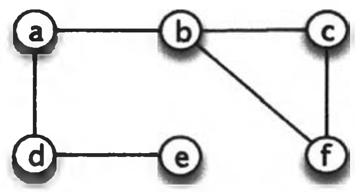

Figure 1.4: A graph with vertices a,b,c,d,e, and f

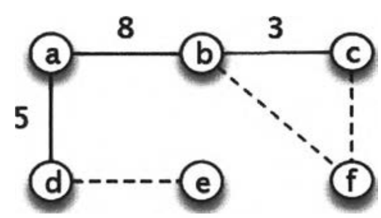

Figure 1.5: The solid-line edges create a path from vertex d to vertex c through vertices a and b.

The existing shortest path problem is to determine the minimum distance between vertices in a graph G consisting of vertex-set *V* and edge-set *E*. Consider the graph representation in Figure 1.4; a circle represents a vertex, and a line connecting two vertices a and b represents an edge between a and b. We denote this edge as (a, b). In Figure 1.4, vertex-set *V* = {a, b, c, d, e, f} and edge-set *E* = {(a, b), (b, c), (b, f), (c, f), (a, d), (d, e)}. A path in *G* from some vertex s to another vertex t is a set of edges that connect s to t through zero or more intermediate vertices. Figure 1.5 represents a path from vertex d to vertex c. For shortest path between some vertex s and another vertex t is a path from s to t with minimum weight (more than one such path might exist). In Figure 1.5, the edges are labeled with numbers, called weights. The weight of an edge can represent any number of characteristics, such as length or transversal time. The weight of a path is the sum of the weights of the edges in that path. The weight function *w(e)* returns the weight of an edge *e*. The weight of the path in Figure 1.5 is sixteen.

The most widely studied variant of the shortest path problem is the Single Source Shortest Path (SSSP) problem. In this problem, a shortest path is found from a source vertex s to all other vertices in the graph. A more interesting subproblem is to find the distance from s to some vertex t. Yet, in general, this problem is no easier than computing the SSSP. The SSSP problem is well studied when edge weights are nonnegative, which is often referred to as the Nonnegative SSSP (NSSSP). In 1959, Dijkstra proposed an elegant algorithm for solving this problem [2]. Little was done to improve time bounds of SSSP algorithms until the 1980s when Fredman and Tarjan introduced the Fibonacci Heap and improved the time bound of Dijkstra's algorithm to $O(m + n \log n)$.

1.3 Research objectives

- To find a new mechanism and a frame structure that perform substantial higher data transfer performance, higher throughput and lower delay, in wireless multi-hop relay network.
- To find a new algorithm and a data structure that perform substantial faster shortest-path query in large weighted-index network than the existing algorithm.

1.4 Scopes of the Study

- To investigate and propose a new methodology of travel-time path selection technique, Adaptive Travel-Time Path Selection (ATTPS), into an adaptive application-level of hierarchical index on road network (HIRN).

- To put forward a new mechanism, network coding-based relay (NC-BR), and frame structure for wireless multi-hop relay network.