Introduction

Space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind union of the two will preserve an independent reality

Albert Einstein (1879-1955)

Space and time are two important aspects of all real world phenomena in the dynamic environment in which we live. Spatial and temporal changes are vital in our daily lives and become an integral part of our life. Extracting knowledge about movement of different type of moving objects enables us to discover mobility patterns that can be explored in applications like service accessibility, mobile marketing, traffic management, security informatics and web click-stream analysis.

1.1 Mobility Management

Any physical object that exists in this universe has a basic characteristic that at any point in time, it is located somewhere. In the world that is evolving rapidly, mobility is an important factor of people’s life. The internet, wireless networks, positioning technology as well as personal devices such as PDAs, cell phones and their related services are getting improved day by day. Over the past few years, rapid advances in miniaturization and personalization of electronic devices have taken place[14Monreals2009], which consequently have resulted in major price reductions. Performance improvements of general computing technologies on the other hand have made it possible to introduce services that previously were even impossible to think of. The ultimate goal of all these advancements is to satisfy the consumer’s rising expectations.

This research is being supported by University Grants Commission(UGC) under Ministry of Human Resources, Government of India under the scheme of Grant to Academic Research for Faculty members in Universities vide order No. 37-633/2009 (SR).
In the coming years, delivering appropriate, timely and personalized services based on the position of mobile consumers will become increasingly important. Such state of the art services can touch upon areas like individual security, e-business, environmental awareness and travel plan.

Despite some success in fulfilling consumer’s requirements, there is still a long way to go and new serious challenges are ahead. One of these challenges is the lack of efficient database to support the mobility scenarios at present. This stems from the fact that existing databases, which is one of the key elements in making more practical and accurate information available, are at their best good in handling static situations, while the concept of mobility brings up a new set of requirements, dealing with dynamic situations.

The central issue in any mobility scenario is the object whose position continuously changes with time, i.e., the moving object. Although the concept of moving object is rather new in the area of spatio-temporal databases, a variety of applications that may benefit from it is enormous. Urban traffic [176Cucchiara2003], especially commuter traffic, and rush hour analysis[177 Giannotti2007]; fleet management and car theft protection; monitoring animal migration; analysis of shopping behavior (in a mall or city centre); patient tracking in a hospital[178 Rebaca2012], inventory tracking, criminals on bail; location-based services[179 Nikos2008], such as tourist information, localized advertising, emergency services; are just a few examples to mention.

1.2 Spatio-Temporal Databases

A database management system (DBMS) is a piece of software that manages a database, a repository of interrelated data items that are often central to the business of an enterprise or institution. A DBMS has the features like data model, query and manipulation language, data independence, data integrity, security and query optimization which allow it to use in variety of applications fields like data mining,
stream processing, information retrieval, workflow and data management \cite{Anisoara2010}. The database managed by standard DBMS normally describes the current state of the world as far as it is known in the database. A change in the current state of the world will be reflected a bit later in some update to the database loosing the previous state. The following limitations of classical databases in standard relational model were identified \cite{Guting2005}.

i. In order to represent geometrical shapes, such as region belonging to a country, hurricane in an area, there is no reasonable way to do this except for very simple objects such as points, for which the coordinates can be represented in numeric attributes.

ii. To represent the development of entities over time such as a time series data, the classical database is not at all useful as it generally reflect the current state of the world and there is no easy way to talk about the past.

iii. In order to represent objects moving around right now or in the past or to forecast its future movement, its movement positions have to be continuously updated, which is not really possible with conventional databases.

The first problem is addressed by the concept of spatial databases which has extended DBMS data models to be able to represent geometrics in a natural way. The implementation of a DBMS needs to be extended by corresponding data structures for geometric shapes, algorithms for performing geometric computations, indexing technique for multidimensional space, and the extensions of the optimizer (transformation rules, cost functions) to map from the query language to the new geometry related components \cite{Guting2005}. Spatial database has so matured that now all the major DBMS vendors (e.g. Oracle, IBM DB2, Informix) offer spatial extensions \cite{Davis1998,Team1999}. Hence the database researchers are trying to build the entire GIS both spatial and attribute data as a layer on top of a DBMS.

Regarding the second limitation, standard DBMS offer very limited support for
managing different database states varying over time[14Guting2005]. This can be achieved by temporal databases [191 Hubert2010] which will integrate time concepts deeply into the DBMS data model and query language. Time can be viewed as discrete or continuous. Discrete models are isomorphic to the natural numbers or integers and continuous models are isomorphic to real numbers. When most people will perceive time as being continuous, for practical reasons temporal database models often use discrete representations of time [180 Barbic1985].

As an alternative to the third limitation stated above is spatio-temporal databases [10 John2003] which will consider the various kinds of data that might be stored in a static (spatial) database and to observe that such data may change over time(temporal)[181 Allen2004]. In this case the database will describe not only the current state of the spatial data, but rather the whole history of this development. Here the attempt is for continuous models which are more appropriate for dealing with moving objects, the major topic of discussion in this thesis. We would like to go back in time to any particular instant and to retrieve the state at that time.

In the real world phenomena, time and space varying in nature must be captured and kept in databases for further processing. In conventional databases, attributes containing temporal or the spatial information are being manipulated solely by application programs[2 Rahim2007], with little help from database management system. A spatio-temporal database is a type of database that supports aspects of both time and space. It offers spatial and temporal data types in its data model and query language [115 Cindy2000, 116 Erwig2000]. A detailed survey on spatio-temporal databases can be found in [118 Abraham 1999, 139 Wolfson1998]. As discussed by Abraham(1999) the forms that spatio-temporal rules may take extensions of their static counterparts and at the same time are uniquely different from them. According to Wolfson(1998), a Database Management System (DBMS) technology provides a potential foundation, upon which to develop spatio-temporal applications. However, DBMS’s are currently not used for this purpose. The reason is that there is a critical set of capabilities that are needed by such applications and
are lacking in existing DBMS’s. These capabilities include location modeling, indexing, uncertainty or imprecision, managing dynamic attributes and spatio-temporal query language.

Thus a spatio-temporal database manages both space and time information. It is an extension of spatial database adding time as one more dimension. A spatio-temporal database embodies spatial, temporal, and spatio-temporal database concepts, and captures spatial and temporal aspects of data and deals with

i. geometry changing over time and/or

ii. location of objects moving over invariant geometry.

Common examples for type (i) data management include:

- A directory of species in a given geographic region, where over a period of time additional species may be introduced or existing species migrate or die out.
- Continuous change in vegetations, human inhabitation in a particular part of the land.

Examples for type (ii) data management are

- Tracking of moving objects, which typically can occupy only a single position at a given time, for example a vehicle moving in a road.
- A database of wireless communication networks eg: mobile adhoc network, which may exist only for a short time span within a geographic region.

This thesis focuses on the spatio-temporal data management in the second perspective which deals with tracking of moving objects which is normally called Moving Object Databases.

1.2.1. Type of Changes in Spatio-Temporal databases

Regarding kinds of change that takes place over time for the spatial data stored in a spatio-temporal database, one has to consider discrete changes and continuous changes. Classical research on spatio-temporal databases has focused on discrete changes for all the spatial entities like point, line, region etc. [Florizzi2000]. In contrast, continuous changes are the topic of this thesis, and this is what is usually
meant by the term “moving object”. Whereas discrete changes occur on any kind of spatial entity, continuous changes seem most relevant for point and region. Hence, a moving point is the basic abstraction of a physical object moving around in the plane or in a higher-dimensional space, for which only the position, but not the extent, is relevant. The moving region abstraction describes an entity in the plane that changes its position as well as its extent and shape as moving region may not only move but also grow and shrink [1Guting2005].

1.3 Moving Object Database

The goal of the research on moving object database is to extend database technology, so that any kind of moving entity can be represented in a database and using powerful query languages, formulate any kind of questions about such movements. Table 1.1 summarizes some examples of moving entities (points) and possible questions(past, present or future oriented information) about them[1Guting2005].

<table>
<thead>
<tr>
<th>Moving Point Entities</th>
<th>Questions</th>
</tr>
</thead>
</table>
| Animals               | • What distance do the birds/whales traverse ? What speed? How often do they stop?  
|                       | • Where are the whales now?  
|                       | • Did their habitats change in the last 20 years? |
| Satellites, Spacecraft, Planets | • Which satellites will get close to the route of this spacecraft within the next 4 hours? |
| Cars: taxi-cabs, trucks | • Which taxi is closest to a passenger request position?  
|                       | • Which routes are used regularly by trucks? |
| Air planes            | • Are two planes heading towards each other (going to crash)?  
|                       | • Were any two planes close to a collision?  
|                       | • Did planes cross the air territory of state X? |
| Ships                 | • Are any ships heading towards shallow areas?  
|                       | • Find “strange” movements of ships indicating illegal dumping of waste |

Table 1.1 Moving Points and Questions
Clearly there exist many kinds of interesting moving entities and one can ask questions about them ranging from simple to very complex. The goal of moving object database research is to design models and languages that allow one to formulate and answer these questions in a simple yet precise way.

(a) Discretely changing point and region

(b) Continuously changing point and region

Fig 1.1 Discrete and Continuous Changes of Point and Region

Table 1.1 emphasize on entities capable of continuous movement. Nevertheless, there exist many applications involving spatial data that change only in discrete steps. To understand the scope of the more traditional spatio-temporal database research let us introduce a classification of time-dependent point and region data. Spatio-temporal data can be viewed in a natural way as being embedded in a space that is the cross-product of the original spatial domain and of time. Here we consider 2D space and restrict attention to a single time dimension, namely valid time. Hence data “live” in a 3D space, as illustrated in Figure 1.1.

Research on moving-object databases has addressed the need for representing movements of objects (i.e. trajectories) in databases, in order to perform ad-hoc querying and analysis [1Guting2005, 3Chamberlain1990].
1.3.1 Classification of Spatio-Temporal data

The spatio-temporal data with respect to their “shape” in the 3D space, can be classified into the following categories[1 Guting2005, 183 Parent1999].

i. *Events in space and time* – *(point, instant)*. Examples are archeological discoveries, plane crashes, volcano eruptions, earthquakes (at a large scale where the duration is not relevant).

ii. *Locations valid for a certain period of time* – *(point, period)*. Examples are: cities built at some time which are still existing or destroyed; construction sites (e.g. of buildings, highways); branches, offices, plants, or stores of a company; coal mines, oil wells, being used for some time; or “immovable’s”, anything that is built at some place and later destroyed/abandoned.

iii. *Set of location events–sequence*(point, instant). Entities of class (i) when viewed collectively. For example, the volcano eruptions of the last year.

iv. *Stepwise constant locations – sequence of (point, period)*. Examples are: the capital of a country; the headquarter of a company; the accommodations of a traveler during a trip; the trip of an email message (assuming transfer times between nodes are zero).

v. *Moving entities – moving point*. Examples are people, planes, cars, etc., see Table 1.1

vi. *Region events in space and time* – *(region, instant)*. E.g., a forest fire at large scale.

vii. *Regions valid for some period of time* – *(region, period)*. For example, the area closed for a certain time after a traffic accident.

viii. *Set of region events – sequence of (region, instant)*. For example, the Olympic games viewed collectively, at a large scale.

ix. *Stepwise constant regions – sequence of (region, period)*. For example, countries, real estate (changes of shape only through legal acts), agricultural land use, etc.
Moving entities with extent – moving region. For example, forests (growth); forest fires at small scale (i.e. we describe the development); people of civilization history.

This research concentrates on data management of moving entities (v) considering the trajectory of object movement.

1.4 Spatio-Temporal Data Mining

We have witnessed an exponential growth of geo-tagged data in the last few years, through the availability of cheap sensor devices, resulting in large digital repository of fine-grained spatial data at small temporal sampling intervals. Therefore, the actual challenge in spatio-temporal analysis is moving from acquiring the right data towards large-scale analysis of the available data\(^{188}\)Kisilevich2010. As this multidimensional data add more complexities in storing, indexing and querying, efficient methodologies have to be derived to manage it using state of the art database technology.

Data mining, the process of finding hidden knowledge within large amounts of data available in databases, has been applied successfully in the past to increase business revenue \(^{5}\)Gidofalvi2007,\(^{166}\)Ian2011. More recently, data mining has been suggested to be useful to derive context and general pattern of user activities in user–friendly Location Based Services. These have made available massive repositories of spatio-temporal data recording human mobile activities, that call for suitable analytical methods, capable of enabling the development of innovative, location-aware applications\(^{4}\)Giannotti2010,\(^{2}\)Yahaya2007. One type of user activity is where and when users were in the past. Data mining methods for such activities have to consider two additional dimensions, namely the spatial, the temporal or jointly the spatio–temporal dimension, and are referred to as spatio–temporal data mining methods\(^{11}\)laube2005 or mobility mining. Due to the unique nature of these two additional dimensions, requiring the management of
multidimensional data, mining of spatio–temporal patterns poses additional challenges.

The GeoPKDD (Geographic Privacy-aware Knowledge Discovery and Delivery), a project[4 Giannotti 2010] funded by European Commission under the Future and Emerging Technologies (FET) program, has been to discover useful knowledge about human movement behavior from mobility data, while preserving the privacy of the people under observation. The various process involved in this project are shown in Figure 1.2. The project proposes the following mobility-related tasks to use spatio-temporal data mining (mobility mining) which aims at improving decision-making at various stages.

- Planning and Monitoring traffic and public transportation systems
- Enforcing security of man and vehicle in transportation networks.
- Forecasting/and simulating traffic-related phenomena.
- Detecting deviations in collective movement behavior
- Innovative info-mobility services
- Localizing new facilities and public services
- Geo-marketing and location-based advertising
- User personalization in Web click stream analysis

1.5 Moving Object in Constrained Networks

As seen in section 1.1, mobility mining usually works on trajectory of moving objects in a free movement space like movements of pedestrian’s, animals in forest, birds or air planes in the sky, ships in the sea and vehicles on roads. A fairly obvious observation is that moving point objects in many cases do not move freely in the 2D plane but rather within spatially embedded networks (on roads, highways – even airplanes normally follow fixed routes). It makes a lot of sense to include such spatial networks in a model and query language for moving objects as well as query processing[1 Guting2005]. The movements can be represented in relation to the road network rather than to the 2D space. For example, instead of describing the
position of a car by geographic coordinates, we can describe it as being a kilometer 260.48 on a particular highway with reference to a point of origin and direction. This has the following advantages [Guting2005, Miller2009].

i) It is easier to formulate queries about relationships between the moving object and the road network, since the road network is a known concept within the database, available as a static (fixed) entity.

ii) A query over moving object on such networks can be evaluated much more efficiently. For example to check which part of a particular highway has been traversed by a given vehicle, in the model with free movement a complex and time-consuming comparison between the geometries of the object and of the highway has to be performed. In the network based model, treating a road as a line, simple comparisons between edge identifiers are sufficient. This simple representation also improves the indexing of trajectories, since fewer entries are needed.

iii) The representation of moving objects becomes much more compact as geometries are factored out (a moving car is being treated as a moving point) with respect to the characteristics of the network. For example, in the standard model a vehicle moving with constant speed along a road

Figure 1.2. Process Involved in GeoPKDD Project
needs a piece of representation with every point of the road. In the network-based model, a single piece of representation suffices as long as the speed does not change.

All the proposed methods in this thesis have taken trajectories of the moving object on constrained network, exploiting the above advantages in data storage, query processing and other such data management tasks.

1.6 Notion of a Trajectory
A trajectory, the basic form of mobility data, is a sequence of time-stamped locations, sampled from the itinerary of a moving object. Vehicles equipped with GPS receivers transmit their positions to a central computer using either radio communication links or mobile phones. At the central site, the data is processed and utilized. To record the movement of an object, its position has to know at all times, i.e., on a continuous basis. However, GPS and telecommunications technologies only allow us to sample an object’s position, i.e., to obtain the position at discrete instances of time, such as every few seconds. A first approach to represent the movements of objects would be to simply store the position samples. This would mean that one could not answer queries about the objects' movements at times in-between those of the sampled positions. Rather, to obtain the entire movement, one has to interpolate, after making the path containing sequences of locations visited, in the form a trajectory.

A database management system and a data warehouse have been designed around this specific form of movement data called trajectory database\textsuperscript{130} \cite{leticia2008}. The design of such trajectory database has been influenced by the research on Moving Object Databases (MOD)\textsuperscript{1} \cite{gutting2005}, which extends the traditional database technology for modeling, indexing and querying trajectory data\textsuperscript{4} \cite{giannotti2010}. The advantage of trajectory based modeling is that it can represent both past and current (as well as anticipated future) positions of moving objects \cite{gutting2005,Moknell2004,Wolfson2000}. Trajectory reconstruction transforms
sequences of raw sample points into meaningful trajectories in accordance with different filters: temporal gaps (nodes are time points), spatial gaps (nodes are location), maximum speed (speed changing points are nodes), tolerance distance (mile stones are nodes), among others. The structure of a trajectory database system warehouse is shown in Figure 1.3.

Generally speaking, trajectories can be classified into two major categories, according to the nature of the underlying spatial object [Frenzos2008]: (i) objects without area represented as moving points, and (ii) objects with area, represented as moving regions; in this case the region extent may also change with time. Among the above two categories, the former has attracted the main part of the research interest, since the majority of the real-world applications, the objects are represented as points, e.g., a fleet management systems which monitor cars in road networks. In this thesis our discussion is confined to trajectories of moving points on constrained networks.

**Figure 1.3** Trajectory database management system and warehouse
TDW- Trajectory Data Warehouse, ETL- Extract Transform, Load Tool
1.6.1 Presentation of a Trajectory

As already discussed a trajectory is a description of physical movement of moving objects changing over time. Database Research has identified the following basic form for the presentation of trajectories\cite{Spaccapietra2008}:

(i) **Raw trajectory** is the recording of the positions of an object at specific space time domain, for a given moving object and a given time interval, it is presented as a sequence of geometric location in 2D spatial system \((x_i, y_i)\) with an additional time dimension \(t\) as shown in Fig 1.4.

![Fig 1.4 Raw Trajectory Representation](image)

(ii) **Structured trajectory** is defined as a raw trajectories structured into segments corresponding to meaningful steps in the trajectory trace (e.g. travel segments) as shown in Fig 1.5.

(iii) **Semantic trajectory** expresses the application oriented meaning using four component (stop, move, begin and end)\cite{VaniaBogorny2008}. Stop, move, begin and end are no more spatio-temporal position, but semantic objects linked to general geographic knowledge and application geographic data as shown in Fig 1.6.
(iv) Other recent approach describes movement patterns in both spatial and temporal contexts based on Region of Interest [141Azedine2012] by defining spatial neighborhood and temporal tolerance as shown in Fig 1.7.

The researcher has taken the raw trajectory format in representing a trajectory in this research since majority of publically available tracking data is in this format where the location is recorded at fixed time stamps. Also it will be suitable for applying dimension reduction and for suitability in exploring applications in web click stream analysis discussed in Chapter 5.
1.6.2 Definition of a Trajectory

A trajectory of a moving object is a polyline in the three-dimensional space, where first two dimensions refer to space and the third dimension to time. It may be represented as a sequence of points \(((x_1,y_1,t_1), (x_2,y_2,t_2), \ldots, (x_n,y_n,t_n))\) with \(t_1 < t_2 < \ldots < t_n\). For a given trajectory \(T_r\), its spatial projection on the xy plane is called the route of \(T_r\). The object at position \((x_i,y_i)\) at time \(t_i\) and during each period \([t_i,t_{i+1}]\), is assumed to move along a shortest straight line from \((x_i,y_i)\) to \((x_{i+1},y_{i+1})\) at a constant speed and that point is obtained by linear interpolation.

A trajectory can be viewed as the history of past movement, the present location of the object and also a motion plan for the future. In this thesis the concentration is more on historical trajectories to analyze (spatio-temporal similarity, clustering) the object movement in the past to predict its movement patterns in the future.

1.7 General problem area of Trajectory data management

Trajectory data are ubiquitous in the real world. Recent progress on satellite, sensor, RFID, video, and wireless technologies has made it possible to systematically track object movements and collect huge amounts of trajectory data. Accordingly, there
is an ever-increasing interest in performing data analysis over trajectory data. The following sections discuss the various kinds of data management over moving object trajectory data.

1.7.1 Modeling and Querying of Moving Object Trajectories

One preliminary function to manage trajectory data is to define frameworks for representing and querying spatio-temporal data in databases. Such frameworks include data models and query languages in order to fully integrate spatio-temporal data in a Relational Database Management System (RDBMS). The proposal in \[^{119}Guting 2000\] presents a set of data types, e.g. moving point or moving region and operations to support querying. A proper integration with existing RDBMS\[^{132}Iulian2009\] is an essential aspect for most industrial or commercial applications to effectively utilize the available indexing and query processing capabilities. Thus, the problem becomes one of recording the position of a moving object across time in a suitable database schema in an RDBMS. The movement of an object may then be represented by a trajectory, or polyline, in the three dimensional \((x,y,t)\) space composed from two spatial dimensions and one time dimension.

Re-construction (modeling) of trajectories from spatio-temporal databases is an important step in further analysis of trajectory data including indexing, querying, finding similarities, clustering etc. The different types of queries \[^{119}Guting 2000\] on trajectory data are:

- **Spatial (range or nearest-neighbor) search**: Find all trajectories that were inside area A at time instant t (or in a time interval) or find the trajectory that was closest to point B at time instant t (or time interval I).

- **Topological / directional search**: Find all trajectories that entered (crossed, left, bypassed, etc.) or were located west (or south) of an area or find all trajectories that crossed (or meet) or were located left of (or right of or in front of) a given trajectory T
1.7.2 Trajectory Similarity Search and Pattern Mining

Generally when the Trajectory Database analytical tools concentrate on presence of moving objects, mobility mining is aimed at analyzing the movement [4Giannotti2010]. A method for mobility data mining tackles two different tasks: first, to define the format of spatio-temporal patterns and models to be extracted from trajectory data, and second, to design and implement efficient algorithms for extracting such patterns and models. Similarity search[120Somayeh2009] is a crucial and fundamental task in data mining.

Trajectory pattern is a novel notion of spatio-temporal pattern, which formalizes the idea of aggregated movement behaviors. A trajectory pattern, as defined in [15Giannotti2007], represents a set of individual trajectories that share the property of visiting the same sequence of places with similar travel times. Therefore, two notions are central: (i) the regions of interest in the given space, and (ii) the typical travel time(or interval of time) of moving objects from region to region. In this approach a trajectory pattern(as shown in Fig 1.8) is a sequence of spatial regions that, on the basis of the source trajectory data, emerge as frequently visited in the order specified by the sequence; in addition, the transition between two consecutive regions in such a sequence is annotated with a typical travel time that, again, emerges from the input trajectories.

![Figure 1.8 A typical Trajectory Pattern](image-url)
For example a pattern may be interpreted as a typical behavior of tourist who rapidly reach a major point of attraction from the point of arrival such as a railway station and spend there about two hours before getting to the adjacent museum. In another example a trajectory pattern may highlight the pedestrian flow of students that reach the university campus from the station: for them, the central bridge over the river is a compulsory passage. It should be observed that a trajectory pattern does not specify any particular route among two consecutive regions, while a typical travel time is specified, which approximates the (similar) travel time of each individual trajectory represented by the pattern. Trajectory based performance analysis has applications in financial sector\cite{Ashok2002} which uses statistical techniques like self Organizing maps, Principal Component Analysis.

1.7.3 Trajectory Clustering

Clustering is one of the general approaches to explore and analyze large amounts of data, since it allows the analyst to consider groups of objects rather than individual objects, which are too many. Clustering associates objects in groups (clusters) such that the objects in each group share some properties that do not hold (or hold much less) for the other objects. Spatial clustering builds clusters from objects being spatially close and/or having similar spatial properties (shapes, spatial relationships among components). Clustering of trajectories\cite{Nanni2002}, implies considering space, time and movement characteristics within a similarity notion: simple distance-based clustering methods are not effective in separating trajectory clusters that exhibit a non convex (non globular) shape, as it often occurs in practice. The goal of trajectory clustering is to find similar movement traces. Many clustering methods have been proposed using different distance measures between trajectories. While most of those studies cluster trajectories as a whole, some of the clustering methods \cite{Li2010} discover similar portions of sub trajectories. Note that a trajectory may have a long and complicated path. Hence, even though two trajectories are similar in some sub-trajectories, they may not be similar as a whole. Discovery of common sub-trajectories is useful, especially if one considers regions of special interest in analysis.
1.7.4 Trajectory Classification and Location Prediction

Predictive models for trajectory data include a classification method [123Mouza2004] for inferring the category of a trajectory, (e.g., the transportation means associated to a trajectory: private car, public transportation, pedestrian, etc.), and a predictor of the next location of a moving object given its past trajectory. There is strong current interest in next location prediction, in that it enables several intelligent location-based services. In the literature, this task is achieved by applying various learning methods to the history of each moving object for the purpose of creating an individual location predictor [142Wolfson2000].

1.7.5 Trajectory Anonymity

Privacy is a big concern, in the context of personal mobility data. ie: Location data allow inferences which may help an attacker to discover private information, such as individual habits and preferences. Hiding explicit identifiers and replacing them with pseudonyms is insufficient to guarantee anonymity [110Francesco 2011], since location represents a property that may allow re-identification: for instance, characteristic locations such as home and work place can be easily uncovered with the use of visual analytics methods, given detailed personal trajectories. Therefore, in all cases when privacy concerns are relevant, the trajectory data cannot be disclosed without appropriate safeguards. Anonymization techniques [121Nikolay 2008,117Noman2009] are data transformations that aim at a double goal: decrease the probability of re-identification below an acceptable threshold, while at the same time maintaining the analytical utility of the data.

1.7.6 Trajectory Uncertainty

Uncertainty is inherent in most spatio-temporal applications due to measurement/digitization errors and missing or incomplete information [10John2003,109Miller2009]. Unless it is captured in the model and query language, the burden of coping with it and reflecting it in the answers to queries will inevitably, and at the same time hopelessly, be left to the user. Assume for instance,
a user with a PDA inquiring about the closest restaurant from his current position in terms of road network distance. Although the user is travelling through the road segment, due to inaccuracy of GPS device, system may fail to recognize this (the location updated by GPS device may be outside the road boundary). Such a situation can be handled by defining an (application dependent) threshold $\theta$, so that if a point is within distance $\theta$ from a road segment, it is assumed to lie on it. Alternatively we can snap the point to the closest road assuming incomplete information, or we can consider it unreachable depending on the application specifications. Similar problems exist for moving object trajectories because while movement is continuous, measurements are discrete.

1.7.7 Visual Analytics

A visual Analytics system will help the analyst to navigate through mobility data and patterns to visually drive the analytical process \cite{Andrienko2008}. The key features include: the visualization of trajectory patterns to support the navigation of the extracted patterns in the spatial and temporal dimensions; the progressive refinement of trajectory clusters, through user-driven exploration and evaluation of the discovered Trajectory clusters \cite{Rinzivillo2008} and the visual exploration of various measures provided by the Trajectory Warehouse \cite{Monreale2009}.

1.8 Motivation towards the Research Problem

This specific research area – spatio-temporal database has ignited the mind of the researcher while preparing for a presentation in the topic ‘GIS & Databases’ to engage a session in the International Workshop on Remote Sensing and GIS’ at Sherubtse College, Royal University of Bhutan on 5th June, 2005. The challenge which came up was how to store both spatial and attribute data together in RDBMS. The enormous potential of this area of research became evident as GIS and location intelligence applications can be the foundation for many location-enabled services that rely on analysis, visualization and dissemination of results for collaborative decision making\cite{Kenneth1995, Michael2010}. The book Moving Object Database by R.H Guting\cite{Guting2005}, gave a glimpse into the latest research in
this field. The final chapter of the book put forward the research opportunities in “objects movements in networks” which incited the curiosity and eventually became instrumental in initializing this study. This area of study is intimately related to human life and could be of significant application if approached scientifically.

1.9 Research Framework and thesis Contributions

The main objective of the conducted research is to provide effective mechanisms that allow Moving Object Databases to efficiently store, query and process historical trajectories using relational database technology utilizing the advantage of binary encoded dimension reduction. This thesis presents 5 works developed for the efficient management of trajectory data on constrained networks as shown in Figure 1.9, representing the overview of the research. As such, the research deals with efficient storage structure representation and query processing of moving object trajectory data, spatio-temporal similarity measure, spatio-temporal clustering, similarity based on sequence alignment and spatio-temporal similarity of web user session trajectories. The summary of contributions of this thesis is described in the following sections, grouped by the respective issue. We are presenting the respective related works in each chapter, due to the variety of issues as discussed above and outlined in Figure 1.9, so as to facilitate a seamless reading of the thesis. The preliminary results have been already published as mentioned in the framework shown in Figure 1.9 and listed in Appendix I.

(i) Modeling and Querying Moving Object Trajectories Using Dimension Reduction

We are dealing with moving point objects on constrained network such as roads, where the travelling route is a known concept. We can then represent movements relative to the network possibly in one spatial dimension (say road location) rather than in the 2D space as two dimensional co-ordinates (x,y). For example, instead of describing the position of a car by geographic coordinates, we can describe it as being kilometer point 180.34 on a particular highway with a reference point and
direction. As a consequence the movement data becomes a set of two-dimensional co-ordinates (location, t) instead of three dimensional (x,y,t).

Figure 1.9 Research Framework with Key Contributions
The advantages of considering such lower dimensional trajectories are the reduced overall size of the data and the lower-dimensional indexing challenge [29Pfoser2003]. Since off-the-shelf database management systems typically do not offer higher dimensional indexing, this reduction in dimensionality allows us to use an RDBMS to store and index trajectories. Moreover, one can see that, given the right circumstances, indexing and querying these dimensionality-reduced trajectories can be more efficient than using a three-dimensional index.

This work has the following objectives.

- To represent road locations and object position as single dimensional binary coded data.
- To keep and structure the moving object trajectory data in relational database scheme and to implement spatio-temporal queries and triggers.
- To evaluate the overhead of the conversion algorithm from two dimension to single dimension.
- To evaluate the efficiency of spatio-temporal queries and triggers over moving object database.
- To design a framework to apply the techniques in Security Informatics Domain.

(ii) Spatio-Temporal Similarity of Network Constrained Moving Object Trajectories

Different types of moving objects share similarities and at the same time they may express differences in terms of their dynamic behavior characterized by the nature of their movement [120Somayeh2009]. Extracting such similarities can significantly contribute to the prediction, modeling and simulation of dynamic phenomena. Therefore, with the development of a quantitative methodology, this research intends to investigate and explore similarities in the dynamics of moving objects by using methods of multi-step query processing architecture.

There are many proposed works [63Tiakas2006, 64Chang2007] concerning trajectory
similarity problem which includes Euclidian, network, time series analysis based measures and concepts known as Points of Interest (POI), Time of Interest (TOI). The broader objective of this work is to demonstrate how these POI and TOI methods could be advantages in security informatics domain suitable to work with road network constrained moving object data, stored at reduced dimension using a binary encoding scheme. The major functional modules in this work are

- Finding spatial similarity based on POI
- Finding Temporal Similarity based on TOI
- Finding Spatio-Temporal Similarity based on POI & TOI
- Evaluating the efficiency and accuracy of the algorithms

(iii) Trajectory Similarity based on Sequence Alignment

The sequence of travel locations visited by a moving object has also to be taken into consideration in evaluating the actual similarity measure. The method involves the following process.

- The structural similarity between locations is obtained by comparing the hierarchical position by finding the binary code differences.
- Finding the sequence similarity by using dynamic programming method.
- Incorporating structural and sequence similarity measures in the existing spatio-temporal similarity process.
- Applying the method by visualizing applications of data management in Vehicular Trajectories.

(iv) Spatio-Temporal Trajectory Clustering

Clustering is one of the most important data analysis methods. The goal is to divide a collection of objects into groups, such that the similarity between objects in the same group is high and objects from different groups are dissimilar. Trajectory clustering has played a crucial role in data analysis since it reveals underlying trends of moving objects. Each clustering process is based on a similarity or dissimilarity measure and a spatio-temporal clustering of trajectory is based on the
spatio-temporal similarity measure already obtained in previous module. The major steps involved in the process are

- Mapping of each trajectory into k-dimensional space as point objects using Fastmap\cite{Faloutsos1995}.
- Clustering the objects using any clustering algorithm
- Experiment the correctness of FastMap by finding the error ratio as the difference between the distance among trajectories before and after the mapping process.

(v) Spatio-Temporal Similarity of Web User Session Trajectories with Applications in Web Usage Mining.

There are a number of resemblances between trajectory of moving object on road network and trajectory generated as a series of web clicks. A web link structure can also be assumed as network constrained since the user clicks can traverse only through predefined paths which is stored as web structure links in web storage space. By analyzing the URL path of each user, we are able to determine paths that are very similar, and therefore effective caching strategies can be applied. The proposed spatio-temporal similarity measure could be applied to web user session trajectories generated from a web user access log. The various functions involved in this work are

- Establishing resemblance between moving object trajectory and web user session trajectory.
- Cleaning Process of Web Access log data and creation of web user session trajectories.
- Variability in applying the similarity measure of moving object trajectory to web user sessions.
- Comparison of results with similar algorithms
- Applying the methods in Security Informatics area of Dark Web Research.
In summary, the main contributions of the research are:


b. Proposing spatio-temporal similarity measure based on Point of Interest and Time of Interest and illustrating its advantage over existing works [104Sajimon2010].


d. The formation of spatio-temporal trajectory cluster set using Multidimensional scaling and Fast Map[103Sajimon2011].

e. Applying the trajectory similarity and clustering methods to click-stream trajectory analysis in Web Usage Mining, visualizing applications in Security Informatics [100Sajimon2011] and Health Informatics[102Sajimon2011].

1.10 Data Sets Used

Throughout this thesis we have experimented the methods proposed with a variety of moving object trajectory datasets which are publically available. Two such data sets which have been used are explained below. The web access log data set of a University web server which has been used for creating the trajectory of web-click sequences is also explained.

(i) Truck Data Set

Trucks dataset [16Theodoridis2007] consists of 276 trajectories of 50 trucks delivering concrete to several construction places around Athens metropolitan area in Greece for 33 distinct days. There are 112,300 position records consisting of the truck identifiers, dates and times, and geographical coordinates. From these raw data, we produced 1100 trajectories by splitting the recordings of a truck in subsets
if there was a temporal gap between two consecutive recordings larger than 15 minutes (a gap that indicates a stop not due to traffic or traffic lights).

The structure of each record is as follows:

\{obj-id, traj_id, time(yyyy-MM-dd HH:mm:ss), x, y\} where (x, y) is in GGRS87 reference system

A sample data set is shown in Table 1.2

<table>
<thead>
<tr>
<th>id</th>
<th>entryid</th>
<th>carid</th>
<th>rdate</th>
<th>rtime</th>
<th>xcoord</th>
<th>ycoord</th>
</tr>
</thead>
<tbody>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:15:59:486253.80;4207588.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:16:29;486261.60;4207543.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:16:59;486292.40;4207562.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:17:29;486289.60;4207473.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:17:59;486226.10;4207270.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:18:29;486225.80;4207094.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:18:59;486233.40;4206933.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:19:29;486276.30;4206825.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:19:59;486292.80;4206720.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0862:1</td>
<td>10/09/2002:09:20:29;486253.30;4206701.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.2.** Sample Truck Data Set

(ii) INFATI Data set

The INFATI data derived from the INFATI Project [17 Jensen2008], an intelligent speed adaptation project carried out by a team of researchers at department of Development and Planning, Aalborg University. For each car that delivered data, the INFATI data contains one file with GPS log data. Table 1.3 describes the important fields of GPS log data in INFATI data set. The Figure 6 uses to represent a pair of x and y coordinates obtained from the GPS receiver, and it uses “+” symbols for positions mapped to the roads. One should note that when the car is near a crossroads, the coordinates are not mapped to the road.

<table>
<thead>
<tr>
<th>id</th>
<th>entryid</th>
<th>carid</th>
<th>rdate</th>
<th>rtime</th>
<th>xcoord</th>
<th>ycoord</th>
</tr>
</thead>
<tbody>
<tr>
<td>991</td>
<td>12091200130310</td>
<td>12</td>
<td>91200</td>
<td>130310</td>
<td>553570</td>
<td>6315889</td>
</tr>
<tr>
<td>992</td>
<td>12091200130311</td>
<td>12</td>
<td>91200</td>
<td>130311</td>
<td>553562</td>
<td>6315863</td>
</tr>
<tr>
<td>993</td>
<td>12091200130312</td>
<td>12</td>
<td>91200</td>
<td>130312</td>
<td>553554</td>
<td>6315836</td>
</tr>
</tbody>
</table>

**Table 1.3.** INFATI data set
Road Network Description

The road network data resides in two files, *road.dat* and *streetId StreetName.txt*. File *road.dat* contains the road geometry, and its format is given in Table 1.4.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>x coord</td>
<td>x coordinate of the road segment</td>
</tr>
<tr>
<td>y coord</td>
<td>y coordinate of the road segment</td>
</tr>
<tr>
<td>street code</td>
<td>Street code of the road to which the road segment belongs</td>
</tr>
<tr>
<td>kmh</td>
<td>Speed allowed on the road segment in kilometers per hour</td>
</tr>
</tbody>
</table>

Table 1.4. Description of File road.dat

A road network is composed of a set of segments. A segment is usually a part of a road that lies in between a pair of consecutive intersections situated along the road. A segment is defined by a sequence of coordinates. Streets are numbered and are composed of several road segments. In file *road.dat*, a segment is thus represented by a set of entries. The value “-9” of attribute *street code* in an entry indicates that the entry contains the last coordinate of a segment. Other values of this attribute...
identify the street to which the segment belongs. A small sample of entries from file road.dat is shown in Table 1.5.

<table>
<thead>
<tr>
<th>x coord</th>
<th>y coord</th>
<th>street code</th>
<th>kmh</th>
</tr>
</thead>
<tbody>
<tr>
<td>55430572</td>
<td>632455870</td>
<td>7486</td>
<td>50</td>
</tr>
<tr>
<td>55430979</td>
<td>632457914</td>
<td>7486</td>
<td>50</td>
</tr>
<tr>
<td>55431749</td>
<td>632458306</td>
<td>7486</td>
<td>50</td>
</tr>
<tr>
<td>55449649</td>
<td>632456885</td>
<td>-9</td>
<td>-</td>
</tr>
<tr>
<td>55419427</td>
<td>632454790</td>
<td>6607</td>
<td>40</td>
</tr>
</tbody>
</table>

**Table 1.5. Entries from File road.dat**

The table contains three segments. The first segment is a polyline described by four coordinates; it has street code 7486. The next two segments are composed of polylines described by three coordinates each, and they belong to the street with code 6607.

File streetId StreetName.dat contains the actual names of the streets. Its structure is described in Table 1.6.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>street code</td>
<td>The code of the street</td>
</tr>
<tr>
<td>street name</td>
<td>The name of the street</td>
</tr>
</tbody>
</table>

**Table 1.6. Description of streetId StreetName.dat**

The road network data was created some time before the GPS log data were collected. As the road network evolves continually, there may be little variation in road network data in which the cars actually traveled during the GPS log data collection. Consequently, there are differences between the roads on which GPS positions were recorded and the digital road network. For practical reason we have ignored this difference for convenience.

**(iii) Web Access Log data set**

Most web servers are configured to record an access log of all client requests for web site content. The access log records information about all the requests and
responses processed by the server. Each line from the access log contains information on a single request for a document. From each log entry, it is possible to determine the name of the host machine making the request, the time that the request was made, and the name of the requested document. The typical syntax of an access log entry is:

```
hostname - - [dd/mm/yyyy:hh:mm:ss tz] document status size
```

The `hostname` is the name or IP address of the machine that generated the request for a document. The following fields ("- -") are usually blank, but some servers record user name information here. The next field indicates the day and time that the request was made, including the time zone (tz). The URL requested is recorded in the `document` field. The `status` field indicates the response code (e.g., Successful, Not Found) for the request. The `size` field indicates the size in bytes of the document returned to the client. A part of the data set is shown in Table 1.7.

```
```

Table 1.7 Web Access Log Data Set

1.11 Organization of the Thesis

The remaining chapters of this thesis are organized as follows:

Chapter 2 presents the modeling of moving object trajectories with the concept of dimension reduction. The techniques of keeping moving object data, road network
data in binary encoding scheme is discussed with its performance evaluation strategies. Query processing and trigger performance over this model and its advantages is evaluated comparing with similar approaches. Illustrations were made in applying the techniques in the domain of security Informatics.

Chapter 3 is discussing on spatio-temporal similarity of moving object trajectories. The broader objective of this work is to demonstrate how POI and TOI based methods could be advantages in security informatics domain suitable to work with road network constrained moving object data, which is stored using binary encoded scheme. This includes algorithms for finding spatial similarity based on POI, temporal Similarity based on TOI, spatio-temporal Similarity based on POI & TOI. It also discusses the usefulness of these algorithms in Security Informatics area for Traffic Security.

Chapter 4 contains an extension work of spatio-temporal similarity measure discussed in Chapter 3 by taking the sequence of travel locations visited by a moving object. The structural similarity of two travel locations has been evaluated by comparing the differences in hierarchical position of locations as the binary code differences. The sequence similarity measure is evaluated using dynamic programming technique.

This chapter also contains an algorithm for trajectory clustering based on multi dimensional scaling approach of Fastmap using spatio-temporal distance measure. The performance of both the algorithms was analyzed using real-life data of small and large trajectory data set.

Chapter 5 emphasizes the fact that there are number of resemblance between trajectory of moving objects on road network and trajectory formed by using web-clicks and hence the proposed spatio-temporal similarity measure could be applied to web user session trajectories generated from web user access log. The various functions involved in the process - Creation of web user session trajectories from web access log, variability in applying the spatio-temporal similarity algorithms to web user sessions, comparison of results with similar algorithms were discussed.
Also the applications of the proposed methods in the Security Informatics area of Dark Web Research have been demonstrated.

Chapter 6 summarizes the research achievements, followed by a critical self-review of the proposed methods and a discussion on the remedies which are assumed as ongoing works. It also provides some guidelines for future research in this field.