TransDec:A Spatiotemporal Query Processing Framework for Transportation Systems (Demo Paper)

Ugur Demiryurek, Farnoush Banaei-Kashani, Cyrus Shahabi

Computer Science Department
University of Southern California Los Angeles, CA 90089-0781
{demiryur, banaeika, shahabi}@usc.edu

***Abstract*— In this paper, we present *TransDec*, an end-to-end data-driven system which enables spatiotemporal queries in transportation systems with dynamic, real-time and historical data. TransDec fuses a variety of real-world spatiotemporal datasets including massive traffic sensor data, trajectory data, transportation network data, and point-of-interest data to create an immersive and realistic virtual model of a transportation system. With TransDec, we address the challenges in visualization, monitoring, querying and analysis of dynamic and large-scale transportation data in both time and space.**

I. INTRODUCTION

The latest developments in wireless technologies as well as the widespread usage of sensors have led to the recent prevalence of Intelligent Transportation Systems (ITS) for realistic and effective monitoring, decision-making, and management of the transportation systems. Considering the large size of the transportation data, variety of the data (different modalities and resolutions), and frequent changes of the data, the integration, visualization, querying and analysis of such data for large-scale real-time systems are intrinsically challenging data management tasks. Due to these challenges, current ITS applications only support limited data monitoring and analysis capabilities.

In this paper, we present and demonstrate a real-world data-driven framework, dubbed TransDec (short for Transportation Decision-Making), which enables real-time visualization, querying, and analysis of dynamic transportation systems. We build TransDec with a three-tier architecture (presentation tier, query-interface tier, and data tier) that allows users to create customized spatiotemporal queries through an interactive web-based map interface. With this architecture, we particularly address the fundamental data management and visualization challenges in 1) effective management of dynamic and large-scale transportation data, and 2) efficient processing of real-time and historical spatiotemporal queries on transportation networks.

TransDec fuses a rich set of real transportation data obtained from RIITS (Regional Integration of Intelligent Transportation Systems) [1] and NAVTEQ™ [2]. The RIITS dataset is collected by various organizations based in Los Angeles County including Caltrans D7, MTA-Metro, LADOT, and CHP. This dataset includes both inventory and real-time data (with update rate as high as every 1 minute) for freeway and arterial congestion, bus location, events, and CCTV snapshots. Moreover, in order to support diverse ITS applications, TransDec contains the transportation network of the entire US, as well as a wide variety of point-of-interest data provided by NAVTEQ.

In addition to offering immersive virtualization of the traffic information and moving assets (e.g., busses, trains), TransDec allows users to issue various real-time and historical spatiotemporal queries about a) traffic at specific segments or sensor stations (with any user-defined level of aggregation at any desired time interval), and b) moving assets and their navigational (e.g., speed, time-to-destination), route, and location-based information.

II. RELATED WORK

Among the numerous ITS applications that have been developed in the last decade, we review two of the most relevant and predominant systems, namely PeMS [3] and ADMS [4].

The freeway Performance Evaluation Monitoring System (PeMS) collects and stores data from loop detectors operated by Caltrans. The main goal of PeMS is to convert this freeway sensor data into graphs that show traffic patterns on highways. PeMS is mainly used to spot bottlenecks, measure the efficiency of highways, and estimate travel times for highway segments. Similarly, ADMS utilizes archived sensor data to measure the performance and operation of Virginia transportation systems. ADMS also provides real-time weather and incident information for specific routes and segments.

Our work is fundamentally different from the aforementioned systems in several ways. First, while the scope of these systems is limited to collection, archival and analysis of the sensor data, with TransDec we fuse the sensor data with various spatiotemporal data (e.g., transportation network data, moving object data, points-of- interest data) to be able to support customized spatiotemporal querying on transportation systems. Second, PeMS and ADMS, for analysis of huge datasets, use traditional Database Management

System (DBMS) specific analytical query processing tools. DBMSs can only support a set of predefined analytical queries. However, with TransDec we utilize more sophisticated query processing techniques that allow for ad hoc and complex analytical queries as well. Finally, TransDec enables various other real ITS applications such as real-time moving asset tracking, route planning and location-based (e.g., k nearest neighbor, range) querying.

III. TransDec Framework

A. Architecture

TransDec adopts a three-tier architecture where presentation, query-interface, and data management tiers are logically separated. With the three-tier architecture of TransDec, the query is initialized at the presentation tier interactively and sent to the query-interface tier where each request is formulated as structured query language before interacting with the data tier. One of the key distinguishing features of TransDec is the provision of an immersive environment that enables users to query the spatiotemporal datasets based on a user-defined area and time interval. Specifically, users can selectively query and display different layers of information on desired regions, and move forward or backward in time for various query types.

To implement the presentation tier, we have integrated a new generation web-based map application, Google™ Map. [5], into TransDec as graphical user interface. As a second choice, we have also developed our proprietary interface (termed Negaah) that provides custom spatiotemporal queries not supported by typical web-based mapping applications.

With the query-interface tier, TransDec allows several independently developed graphical user interfaces (GUI) to interact with our spatiotemporal data tier transparently. Our query-interface tier (developed in Java, XML, and SOAP) offers a universal standard for specifying the type of query (e.g., shortest path, range aggregate, etc.) and its parameters, as well as the returned results.

The data repository of TransDec is a spatiotemporal database management system built on Oracle™. 10g [6]. Our data repository stores a variety of both dynamic (frequently changing) and static datasets such as highway sensor data, road network information (i.e., vector data), moving object trajectory data, point-of-interest data (e.g., hospitals, restaurants), terrain data, satellite and aerial imagery, and raster maps. Most of the data stored in the repository are labeled by both space and time to allow for a wide range of spatiotemporal queries. In addition, this tier includes a data fusion engine which continuously acquires sensor readings from RIITS web-services and stores them into the data repository. In the rest of this section, we describe the major static and dynamic multimodal data components of TransDec.[1]

1) Sensor Data: TransDec, through RIITS, acquires traffic sensor data from approximately 1500 sensors located on 18 highways (covering 1183 miles) at the boundaries of Los Angeles County (LA). The arrival rate of the data from each sensor is 1 reading/sensor/min. The storage space required for this streamed dataset is approximately 350 MB/day (120 GB/year) without indexing overheads. Currently, our database consists of about 500 million sensor reading representing traffic patterns on the highway segments of LA. We are currently developing necessary modules to collect the sensor readings from major artillery roads in LA and closed-circuit television (CCTV) data from the highways.

[1]Information on these datasets and web-services is publicly available [7], [8].

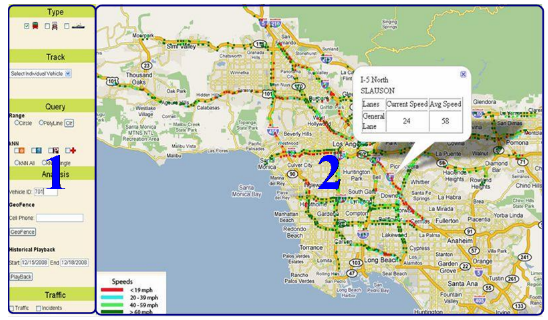


Fig. 1. TransDec's graphical user interface

2) Transportation Network Data: TransDec contains a wide range of transportation network data provided by NAVTEQ including highways, major and secondary roads, streets, railroads, bridges, etc., for the entire US. Each network segment is represented in the vector data format and described by more than 20 attributes such as direction, speed limit, zip code, paved, etc. In addition, TransDec includes more than 15 layers of point data (e.g., bus/train stations, restaurants, hospitals, etc.) also provided by NAVTEQ.

3) Trajectory Data: Another spatiotemporal dataset of TransDec consists of the trajectory information collected from moving objects whose location in the space changes over time. Currently, TransDec collects live location data from GPS equipped USC trams and student cell phones [9].

B. Spatiotemporal Query Processing

We categorize the current queries adopted by TransDec into following four types: 1) Monitoring queries on streaming data, 2) Analysis and mining queries on historical data, 3) Route planning queries, and 4) Location-based queries (e.g., nearest neighbor and range queries). TransDec's graphical user interface (Frame 1 in Figure 1) allows users to formulate these queries by determining the three main component, namely Data/Object Type, Query Type, Spatial and/or Temporal Bounds. Subsequently, TransDec visualizes the query results and allows user to further interact with the interface and/or formulate more queries (Frame 2 in Figure 1) on the selected objects. In particular, every object that is visualized (2 in Figure 1) in the system is selectable to retrieve more information and/or can be used in new queries. We explain each query type in the rest of this section.

1) Monitoring Queries on Streaming Data: With monitoring queries on streaming data, the data is in the form of constantly received data streams rather than finite stored datasets,

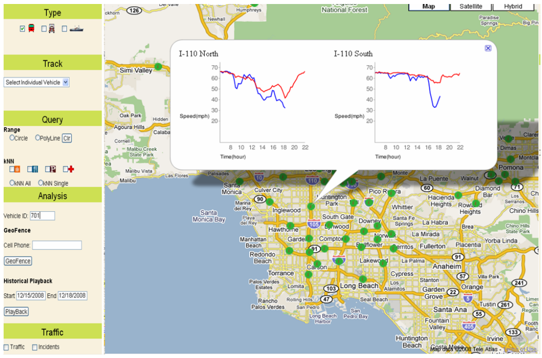


Fig. 2. An OLAP query on sensor data

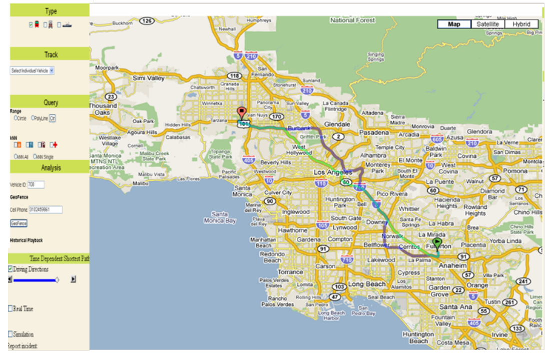


Fig. 3. Time-dependent route planning

and the queries are continuous as opposed to one-time queries. For example, consider a query that continuously reports the speed and occupancy information from each highway sensor every ten seconds. With another example, consider a query that tracks the real-time geographical location, speed and time-to-destination information of a moving object. TransDec currently includes two types of monitoring queries on streaming data collected from a) highway traffic sensors in LA County and b) GPS enabled moving objects (USC trams and student cell phones). In order to keep the results up-to-date for continuous monitoring, we update the result sets incrementally with the arrival of new data tuples rather than querying the entire datasets in defined intervals.

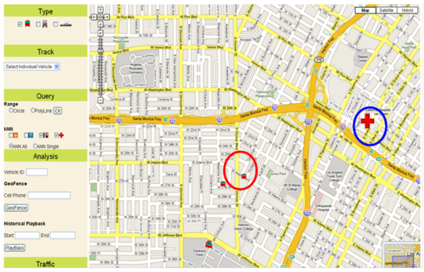
2) Analysis and Mining Queries on Historical Data: In TransDec, the goal of analysis and mining queries is to discover useful knowledge and extract patterns from the historical data. To exemplify, consider a query that reports the average speed of a segment (e.g., on I-5 from post mile 293 to 300) within last five Mondays between 8:00am to 8:30am. To perform such analysis on the historical datasets, we use two techniques. While we use online analytical query processing (OLAP) solutions for relatively small datasets, we utilize wavelet based analytical query processing techniques (developed by our Lab) [10], [11] for larger datasets. Figure 2 depicts two (North and South bound) speed graphs that are drawn based on the results of a spatiotemporal aggregate query on historical traffic sensor data. The red plot represents the historical average of a particular highway segment within a particular day, and the blue plot shows the real-time speed reading on the same segment.

3) Route Planning Queries: The traditional route planning approaches (i.e., minimum travel time path between the source and destination) been studied extensively in the past. With TransDec, we introduce a new class of route planning query namely, time-dependent route planning (TDRP). Unlike the existing studies, during the shortest path computation, we assume the edge costs of the network is function of the time (time-dependent) rather than fixed. This assumption suggests that the arrival time on a segment entry determines the travel-time on that segment. Hence, to compute the fastest path from a source to a destination, all combinations of arrival times at all possible segment entrances must be considered that could exponentially grow the search space. Subsequently, we show that the fastest path from a source to a destination may vary significantly depending on the time of the day. Figure 3 illustrates an example of TDRP where blue route is computed based on the time-dependent edge travel times and green route is computed with fixed edge travel times for a user-defined departure time from the start point (green icon) to the destination (red icon). The blue slider on the left menu enables users to move forward or backward in time by allowing them to select various departure times from the start point. In a real-world scenario, TDRP application can be used to alert the travelers about the best departure time and the realistic expected travel time for a given path.

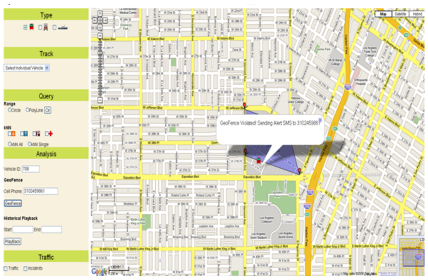
4) Location-Based Queries: The location-based queries adopted by TransDec are spatial queries, including k nearest-neighbor (k-NN) and range queries that look for desired point-of-interest (POI) to a referred object or location. Location-based queries differ from conventional queries as they a) use geometric data types such as lines and polygons, and b) consider the spatial relationship between these geometric data types. Therefore, implementing location-based queries require different data and index structures which support multidimensional selection and join operations by spatial criteria. The main challenges in processing location based queries are as follows: a) the mobility of query points and/or POI, b) fast road network distance (shortest path) calculation, and c) efficient update of dynamic index structures.

A k-NN query finds the k point-of-interest (ie: restaurant, taxis) that is closest (in network distance) to a given query point. With k-NN queries, both the POI and the query points could be either static or dynamic (moving), which results into combination of four cases. To implement k-NN queries, depending on the movement types, we utilize several query optimization and indexing techniques (e.g., Voronoi, grid, R-Tree) which we have developed [12], [13]. In addition, with one of our recent study [14], we have developed a time-dependent k-NN search algorithm in spatial network databases. Figure 4(a) illustrates an example of k-NN query where the nearest hospital (circled in blue) to a moving vehicle (circled in red). As the vehicle moves along its path, the query result (the nearest hospital) is continuously updated.

A range query is a spatial query which returns all POIs



(a) k-NN query



(b) GeoFence query

Fig. 4. Location-based queries

inside a user-specified bounding box such as circle or polygon. For example, in TransDec, a user may draw a circle with a center any point on the map interface and ask the system to return buses within two miles radius of that center. An advanced range query type supported by TransDec is GeoFence query. The GeoFence query allows users to create a virtual monitoring boundary on a geographic area on the map. When that boundary is crossed by a moving object, the intrusion is recognized as an event and the user is notified by a SMS text or email message. Figure 4(b) shows an example of GeoFence application where TransDec sends a SMS message to the user in the event of a USC tram leaves the school boundaries (blue region in Figure 4(b)). It is important to note that there are several commercial applications of GeoFence. All these applications [15], [16], [17] make the simplifying assumption that the distance between the objects/points is Euclidean distance during the query processing. In real-world scenarios, however, the objects move in spatial networks, where the distance between the objects is the length of the shortest path (i.e., network distance) connecting them. Therefore, TransDec considers the network distance during the query processing hence can easily be applied to real-world road network applications.

With TransDec, it is also possible to query the trajectory of the moving objects. The trajectory-based queries are usually combined with k-NN and range queries hence inheriting the similar challenges. To exemplify, consider the queries such as "show me the trajectory of Bus-A between 8:30am and 9:30am today", and "find all the buses that were inside the boundaries of Los Angeles between 9-12am yesterday".

IV. Demonstration

The TransDec demonstration presents live and recorded interactive sessions showcasing the selected examples of the aforementioned query types interest to various domains including transportation, emergency response, location based services. Multiple geographical areas are featured including the USC Campus and Los Angeles County.

V. Conclusion and Future Work

In this paper, we introduced TransDec, a real-world data-driven system that enables interactive and extensive querying of transportation related spatiotemporal datasets including traffic sensor data, trajectory data, transportation network data, and point-of-interest data. We explained the design and implementation details of TransDec's three-tier architecture. We also introduced a set of advanced spatiotemporal queries supported by TransDec. We intend to pursue this work in three directions. First, we plan to extend the capabilities of the query-interface tier to support more complex ad hoc analytical queries. Second, we intend to extend the set of queries supported by TransDec to a complete minimum set that allows for formulation of generic decision-making queries in all typical ITS applications. Finally, we plan to improve the presentation tier of TransDec to port it to mobile computing devices.

VI. Acknowledgments

This research has been funded in part by NSF grants IIS-0238560 (PECASE), IIS-0534761, and CNS-0831505 (CyberTrust),the NSF Center for Embedded Networked Sensing (CCR-0120778) and in part from the METRANS Transportation Center, under grants from USDOT and Caltrans. Any opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] RIITS System. http://www.riits.net/. Last visited December 25, 2008.
[2] Navteq. http://www.navteq.com. Last visited June 17, 2009.
[3] PeMS: Freeway Performance Measurement System. https://pems.eecs.berkeley.edu/. Last visited May 15, 2009.
[4] ADMS Smart Travel Lab. http://cts.virginia.edu/stl-adms.htm/. Last visited May 20, 2009.
[5] Google Maps. http://code.google.com/apis/maps/. Last visited June 2, 2009.
[6] Oracle. http://www.oracle.com. Last visited December 1, 2008.
[7] InfoLab Data. http://infolab.usc.edu/database.htm. Last visited June 15, 2009.
[8] Infolab Web Services. http://infolab.usc.edu/web-services.htm. Last visited June 15, 2009.
[9] GeoSim Project. http://infolab.usc.edu/projects/GeoSIM/. Last visited June 21, 2009.
[10] C. Shahabi. Making sense of data overload:an innovative approach to progressive data analysis. In *Microsoft Research. http://research.microsoft.com/en-us/collaboration/papers/usc.pdf.*
[11] M. Jahangiri and C. Shahabi. Plot query processing with wavelets. In *SSDBM*, 2008.
[12] M. Kolahdouzan and C.Shahabi. Voronoi-based k nearest neighbor search for spatial network databases. In *VLDB*, 2004.
[13] U.Demiryurek, F. Banaei-Kashani, and C. Shahabi. Efficient continuous nearest neighbor query in spatial networks using euclidean restriction. In *SSTD*, 2008.
[14] U.Demiryurek, F. Banaei-Kashani, and C. Shahabi. Time-dependent k nearest neighbor search in spatial networks. Submitted for review.
[15] OnStar by GM. http://www.onstar.com. Last visited June 20, 2009.
[16] Loopt. http://www.loopt.com. Last visited June 20, 2009.
[17] Telenav. http://www.telenav.com/. Last visited June 20, 2009.