Program Name:
   NGA Academic Research Program (NARP)

Proposal Category:
   NURI

BAA Number:
   BAA # HM0177-12-BAA-0001

Title
   Identifying and Analyzing Patterns of Evasion

PI:
   Prof. Shashi Shekhar
   Department of Computer Science and Engineering
   University of Minnesota - Twin Cities

Abstract:
   We propose to develop space-time aware methods for modeling patterns of evasive behavior by insurgents and other security targets. These targets increasingly employ techniques to mask their movements and locations. Denial, deception and evasion techniques skew data collection and hinder traditional data mining techniques. We propose an overarching framework to identify and analyze these denial and deception instances. First, we propose a method to distinguish between evasive and non-evasive behaving groups by quantifying the space-time entropy or predictability of individuals’ behavior. Second, we will identify “blackholes,” areas where no target movement is observed, despite predictions that such movement would occur. Third, since conventional data mining techniques cannot be applied in areas lacking reported observations (blackholes), we look to theoretical understanding of human behavior to help generate hypotheses about target location and travel routes. Specifically, we apply routine activity theory, a well-known theory used by environmental criminologists, which holds that individuals typically follow set patterns in their daily lives. To help quantify this behavior, we look to the physical science concept of return periods, the notion of certain events happening with a given occurrence rate (e.g., 100 year earthquake of certain magnitude). We utilize return periods to estimate a target’s schedule to aid in interception and surveillance.
1. Program Name:
   NGA Academic Research Program (NARP)

2. Proposal Category:
   NURI

3. BAA Number:
   BAA # HM0177-12-BAA-0001

4. Title
   Identifying and Analyzing Patterns of Evasion

5. Research Area:
   Understanding Human Activities, Predictive Intelligence

6. Principal Investigator:
   Shashi Shekar, Professor, University of Minnesota, Department of Computer Science and Engineering, 4-192 EE/CS Building 200 Union Street SE, Minneapolis, MN 55455, Phone: (612) 625-4012 Fax: (612) 625-0572, Email: shekhar@cs.umn.edu

   Alternate contact information:
   Claudette Caparelli, Grants Management Coordinator
   Phone: 510-734-2453, Fax: 612-625-0572 Email: caparell@cs.umn.edu, grants@cs.umn.edu

7. Signature of Principal Investigator. ___________________________ Date _______________

8. Applicant Institution
   Regents of the University of Minnesota, DUNS 555917996, TIN 416007513

9. Administrative Officer
   Kevin McKoskey, Senior Associate Director, Sponsored Projects Administration
   University of Minnesota, 200 Oak St SE, Suite 450, Minneapolis, MN 55455-2070
   Phone: 612-624-5066, Fax: 612-624-4843, Email: awards@umn.edu

   Alternate contact information:
   Jeff Zytkovicz, Grants Administrator, Sponsored Projects Administration
   University of Minnesota, 200 Oak St SE, Suite 450, Minneapolis, MN 55455-2070
   Phone: 612-624-0534, Fax: 612-624-4843, Email: zytko004@umn.edu

10. Signature of Administrative Officer. ___________________________ Date _______________
1 Objectives

We propose to develop space-time aware methods for modeling patterns of evasive behavior by insurgents and other security targets. These targets increasingly employ techniques to mask their movements and locations [16]. For example, Osama bin Laden’s courier did not use his phone within many miles of the compound Osama was located in [15]. These denial, deception and evasion techniques, where targets purposely mask their movements and locations, result in skewed collection datasets that do not accurately represent the ground-truth.

Two lines of research highly relevant to modeling patterns of evasion are: 1) methods for mining trajectory patterns such as hotspots, anomalies, and clusters [12, 17, 18] and 2) methods for quantifying the predictability of movement [4–6, 8, 13]. In its current form, however, current approaches have significant limitations in geo-intelligence settings. The pattern mining methods are not designed to work on non-representative datasets skewed by denial and deception events. Movement predictability algorithms cannot account for evasive behavior, such as the turning off of phones as targets approach a specific location.

We address these limitations with two sets of novel ideas: First, we propose a method to distinguish between evasive and non-evasive behaving groups by quantifying the space-time entropy or predictability of individuals’ behavior. Second, we identify “blackholes,” areas where no target movement is observed, despite predictions that such movement would occur. Third, since conventional data mining techniques cannot be applied in areas lacking reported observations (blackholes), we look to theoretical understanding of human behavior to help generate hypotheses about target location and travel routes. Specifically, we apply routine activity theory, a well-known theory used by environmental criminologists, which holds that individuals typically follow set patterns in their daily lives. To help quantify this behavior, we look to the physical science concept of return periods, the notion of certain events happening with a given occurrence rate (e.g., 100 year earthquake of certain magnitude). We utilize return periods to estimate a target’s schedule to aid in interception and surveillance.

2 Anticipated Results

Anticipated results of this project include new mathematical models, computer algorithms, data analysis methods, and analytical tools. Specifically, we anticipate the following results:

- **Mathematical models:** We anticipate developing a graph-based spatio-temporal data model to represent the fluid and ad-hoc nature of complex networks. We expect that the proposed data model will provide an efficient way to represent return periods.

- **Computer algorithms:** We expect to design new scalable data mining algorithms for finding each portion of the proposal: space-time entropy, blackholes, theory-based movement predictions and return periods. These algorithms will need to scale to large datasets of billions of GPS records.

- **Data analysis methods:** We will develop new interest measures to quantify the proposed patterns-of-life. This measures will consider computational scalability to maintain applicability to spatial big data.

- **Analytical tools:** We expect to develop new software and extend the capabilities of existing software such as Oracle Spatial to facilitate a complete workflow from dataset input to actionable information.
3 Applicability

According to the Washington Post, Air Force Col. Eric J. Holdaway reported, “‘One of the target areas is the eastern border of Afghanistan’, along which Holdaway said he has learned there were ‘278 distinct mountain passes’ to Pakistan. Tribes live on both sides of the border, which to them ‘is more theoretical than anything’ [14].” We hypothesize that non-evasive tribal groups operate (e.g., villagers, nomads, traders) with different patterns-of-life, or different movement patterns, than evasive terrorists due to standard advice to all combatants: “Predictability kills”, Johnson said. ‘Travel down a road, and you’re very predictable. You’re only going one of two directions. That makes it very easy for our enemy to place IEDs (improvised explosive devices) or (set up) ambushes that kill our soldiers [16].” For example, in Figure 1, soldiers are advised to avoid geo-tagging, which can reveal their location to adversaries [1].

Areas of evasion are well known in routine activity theory, where criminal offenders typically do not travel far from known areas as they may be less familiar with their surroundings, having a greater chance of standing out, and a harder time getting away, etc. At the same time offenders also do not want to commit crime too close to home, as they might be recognized. The interplay between these two scenarios produces a buffer around an offender’s residence, referred to as a ‘donut hole’ in geographic profiling [10]. A similar situation may unfold when modeling space-time patterns-of-life of a terrorist or drug dealer, where activities may only be observed outside a certain buffer around their house. A famous case of terrorists creating buffers around their living compounds, Osama bin Laden, had people sometimes driving “90 minutes away before placing a battery in a cellphone” [15]. In Figure 2, we illustrate the movement patterns of tribes (Figure 2(a)) and insurgents (Figure 2(b)). The trajectories are continuous in Figure 2(a), but not in Figure 2(b), where a denial practice (e.g., turning off cellphone) has resulted in a loss of tracking in the center square. We believe that routine activity theory as used in criminology has applications for modeling space-time patterns of terrorists across the globe. The routine nature of their behavior can be exploited by the proposed approach to identify and provide actionable intelligence about locations and schedules of target individuals.

4 Approach

We propose a two-phase framework, shown in Figure 3, for mining patterns of evasion in trajectory datasets. The first phase identifies patterns of evasion. First, we propose a method to distinguish between evasive and non-evasive behaving targets by quantifying the space-time entropy (predictability) of individuals’ movement. A complementary algorithm identifies “blackholes,” areas where no target movement is observed, despite predictions that such movement would occur. A list of targets found to have patterns of evasion are generated. Further analysis can help in spatio-
temporal prioritization of potential sighting and interception places and time-slots of targets on
the list. We will investigate two approaches: (1) movement hypotheses for predicting movement
in denied areas and (2) return period and schedule determination. This second phase leverages
physical science theories, (e.g., continuity of movement) and social science theories (e.g., routine
activity theory) suggesting that individuals typically follow set patterns in their daily lives.

Figure 3: Analyst Workflow using Proposed Approach. Ovals represent processes, while boxes
represent data.

5 Science
The following section details the four proposed ideas: Space-time Entropy, Blackholes, Theory-
based Movement Hypothesis, and Return Periods.

5.1 Space-Time Entropy Discrimination
We intend to research entropy measures for use in identifying target groups based on historical
movement patterns. Entropy is a measure of unpredictability, so a space-time entropy measure
applied to a target would quantify the unpredictability of a target’s movement. Shannon’s Diversity
Index\(^1\) (SDI) is a commonly used measure of entropy, essentially quantifying the predictability
of an individual or event. As detailed in Table 1, the insurgent in Figure 2(b) has a higher SDI value
than the tribesmen in Figure 2(a) (i.e., due to having more known locations to visit, assuming a uni-
form probability of a target being at any one of the given locations). However, there is a substantial
semantic gap between Shanon’s Diversity Index and the notion of spatio-temporal predictability.

\[^1\]Shannon’s Diversity Index: \(H(X) = -\sum_{x=1}^{\text{places}} p(x) \log_2 p(x)\), where \(p(x)\) is the probability of an event \(x \in [1,2,...,\text{places}]\)
### Table 1: Shannon’s Diversity Index for Figure 2.

<table>
<thead>
<tr>
<th>Group</th>
<th>Places</th>
<th>( p(x) )</th>
<th>Entropy (Shannon’s Index)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tribesmen</td>
<td>3</td>
<td>( \frac{1}{3} )</td>
<td>(- \sum_{x=1}^{3} p(x) \log_2 p(x) = - \frac{1}{3} \log_2 \frac{1}{3} = \log_2 3 = 1.58)</td>
</tr>
<tr>
<td>Insurgents</td>
<td>6</td>
<td>( \frac{1}{6} )</td>
<td>(- \sum_{x=1}^{6} p(x) \log_2 p(x) = - \frac{1}{6} \log_2 \frac{1}{6} = \log_2 6 = 2.58)</td>
</tr>
</tbody>
</table>

For example, while SDI works in the simple case illustrated in Figure 2 (i.e. e.g., uniform distribution), it does not account for the properties of spatio-temporal datasets such as the spatial spread and the spatial clumpiness of the data. In Figure 4(a), SDI will return the same value for the two groups (stars and squares) even though the squares have a more concentrated pattern which makes this pattern more spatially predictable. A similar problem is inherent in the partitioning of space using a uniform grid or raster. Depending on the partition size, SDI scores change even if the underlying data does not. This scenario is depicted in Figure 4(b), where SDI values differ from Figure 4(a) even though the underlying data has not changed. Closing this semantic gap requires conceptual and computational innovations.

**Task S1: Conceptual Modeling to of Space-Time Entropy:** To address the limitations of SDI illustrated in Figure 4, we propose to develop a conceptual model to represent the spatio-temporal predictability of targets’ movement across a space. We propose to investigate spatial concepts such as: range, auto-correlation, and resolution. Spatial range distinguishes the size of an area of operation. For example, Africa has a larger range than Mali. Spatial auto-correlation may indicate the existence of clustered data points which makes targets more predictable. Spatial resolution may account for the effect of fine-resolution versus coarse-resolution of space partitioning on predictability. We propose to investigate the following questions: Are range, auto-correlation, and resolution necessary and sufficient to close the semantic gap between SDI and spatio-temporal predictability? If not, then what set of spatial and spatio-temporal concepts are needed?

**Task S2: Designing Quantitative Interest Measures for Space-Time Entropy:** Given a set of concepts modeling space-time entropy (from Task S1), the goal of this task is to design a set of quantitative interest measures. This task is challenging due to the trade-off between the need for a statistical interpretation to evaluate the confidence of the proposed measures and the need for efficient algorithms designed to compute these measures. We propose to investigate current quantitative measures of range (e.g., Minimum Bounding Rectangle (MOBR)), auto-correlation (e.g., Moran’s I, Ripley’s K), and resolution for statistical interpretation and computational scalability.

**Task S3: Scalable Algorithms for Computing Space-Time Entropy in Spatial Big Data:** Given
the quantitative interest measures defined in Task S2 and a large set of GPS trajectories, the objective of this task is to develop efficient algorithms for computing these measures. The computational complexity inherent in evaluating interest measures such as auto-correlation represents a big challenge for designing scalable algorithms. We plan to investigate data filtering and refinement techniques for selecting the most relevant subset of input trajectories for which the values of the interest measures are likely to be more significant.

**Task S4: Validation:** The proposed algorithms will be evaluated using theoretical and experimental methodologies. Theoretical methods, e.g., proofs, will be used to characterize correctness and completeness. Experiments with synthetic and real datasets will be used to evaluate the computational efficiency of proposed algorithms under different combinations of parameters (e.g., dataset size, spatial footprint).

### 5.2 Blackhole and Patterns of Evasion Detection:

We use the term ‘blackhole’ to describe a common denial phenomenon in surveillance: targets that avoid use of electronic devices in an area of interest. One way to detect these areas is to identify significant mismatches between expected observations and recorded observations, assuming persistent surveillance. In Figure 2(b), we illustrate cell phone reports from a group of insurgents within 6 camps. However, there are no reports in the middle of the picture despite the expectation of continuity in trajectories moving across the space. There is initial movement towards the central area, but no recorded observations once the center cell is reached. This type of pattern is indicative of a blackhole. We propose to automatically identify blackholes by building activity patterns for each target, extrapolating potential paths between events using expectations of continuity in movement and aggregating them over time. In Figure 5(a), we extrapolate (using shortest-path computation) each trip’s potential route, and highlight an area of interest that appears to be devoid of reports but consists of multiple expected crossing movement paths where observations are expected due to continuity assumptions. The green segments in the figure represent fast travel speed, whereas the red indicates a long duration between recorded observations, indicating a gap in coverage.

Related work in hotspot and anomaly detection for movement analysis identify clusters of objects (moving or static) [12], increased activity on road networks [9], and other directly observable phenomena [7]. The approach we propose makes it possible to consider the lack of observable data points, or essentially...
the opposite of a hotspot.

**Task B1: Model Blackholes:** Lack of recorded observations of a target despite expectation may indicate areas of interest. Using background information (e.g., other cellphone users), we can model and estimate the expected continuity of observation through denied areas. We will create a conceptual model defining classes of blackhole events (e.g., known cell-phone dead spot or abnormal area). The sub-types will differentiate between total population evasion (cell tower stops functioning) and partial evasion (some number of phones evading a certain area). We will be collaborating with security analysts on this task to ensure grounding with real-world events.

**Task B2: Develop Interest Measures for Blackhole Evasion:** Quantitative interest measures of evasion will be needed to identify various blackhole events. Ideally the measure can not only characterize a type of blackhole event, but also lead to computationally efficient algorithms based on properties like anti-monotonicity.

**Task B3: Design Algorithms for Identifying Blackholes:** We will design new scalable blackhole event mining algorithms, using quick pruning of irrelevant candidates using properties of the quantitative interest measures. We propose efficient space-time prisms to measure expected movement in observation gaps to estimate continuity of movement to identify blackhole areas.

**Task B4: Validation:** The algorithms will be evaluated using theoretical and experimental methodologies. Theoretical methods, e.g., proofs, will be used to characterize correctness and completeness. Experiments with synthetic and real datasets will be used to evaluate the computational efficient of proposed algorithms under different combinations of parameters (e.g., dataset size, spatial footprint).

### 5.3 Theory-based Movement Hypothesis Generation:

Due to the lack of target observations inside identified blackholes, it is difficult to use traditional data mining methods to generate hypotheses regarding target movement. We need a means to predict behavior that does not rely on direct observations of target movement. Further, it needs to be able to leverage features that are observable inside the blackhole, such as transportation networks and points-of-interest. For example, in Figure 5(b), we see tracks ending and starting (represented by circle endpoints) around the outskirts of a small city. By calculating likely routes we could connect the dots and fill in the missing pieces. Then we can aggregate interpolation across multiple tracks to hypothesize likely routes, main corridors and key locations that may be frequented by targets practicing denial. Our early work [2] in this area aggregated tracks within cities based on temporal feasibility, and required precise timing. The proposed work will be more robust to temporal differences but may require more temporal aggregation to accumulate enough tracks for precise location hypothesis generation.

Once blackhole areas have been identified, routes and locations of interest can be hypothesized based on recorded observations using existing theories in relevant fields (e.g., routine activity theory in environmental criminology, kinematics in physics), the underlying transportation network and expectations such as continuity of trajectories over space-time.

**Task T1: Develop a Conceptual Model for Network-based Movement Hypotheses Generation:** Even lacking observations in blackhole areas, we can utilize the underlying spatial features to generate hypotheses about movement. We propose to develop a conceptual model for hypothesized (multi-source, multi-destination) movement generation based on trajectory fragments (tracklets).
Table 2: Arrival Count at Place and Time over 30 Days *(Return Period in Italics)*

<table>
<thead>
<tr>
<th></th>
<th>Morning</th>
<th>Afternoon</th>
<th>Evening</th>
<th>Night</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td>10 <em>(3)</em></td>
<td>2 <em>(15)</em></td>
<td>15 <em>(2)</em></td>
<td>19 <em>(1.5)</em></td>
<td>44</td>
</tr>
<tr>
<td>Work</td>
<td>19 <em>(1.5)</em></td>
<td>20 <em>(1.5)</em></td>
<td>10 <em>(3)</em></td>
<td>1 <em>(30)</em></td>
<td>50</td>
</tr>
<tr>
<td>Club</td>
<td>4 <em>(7.5)</em></td>
<td>5 <em>(6)</em></td>
<td>4 <em>(7.5)</em></td>
<td>0 <em>(∞)</em></td>
<td>15</td>
</tr>
<tr>
<td>Farm</td>
<td>0 <em>(∞)</em></td>
<td>0 <em>(∞)</em></td>
<td>1 <em>(30)</em></td>
<td>0 <em>(∞)</em></td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>20</td>
<td>110</td>
</tr>
</tbody>
</table>

**Task T2: Design Graph-based Data Structures and Operators for Movement Hypothesis:** Tracklets, being only a small part of an overall trajectory, making it difficult to locate the actual starting location and destination. By hyper-joining tracklets, we can use aggregate enumeration to identify potential source and destination connections. Enumerating these movement hypotheses will require efficient data structures and operators for quickly mining interesting paths.

**Task T3: Formulate Algorithms for Generating Movement Hypothesis of Targets:** Due to the large search space of hypothesized movement from tracklets, scalable algorithms are required to quickly find likely paths of specific individuals. In addition, found paths will need to retain the original tracklets to support analyst confirmation.

**Task T4: Validation:** The proposed algorithms will be evaluated using theoretical and experimental methodologies. Theoretical methods, e.g., proofs, will be used to characterize correctness and completeness. Experiments with synthetic and real datasets will be used to evaluate the computational efficient of proposed algorithms under different combinations of parameters (e.g., dataset size, spatial footprint).

### 5.4 Return Period in Movement Datasets

Routine Activity Theory and Environmental Criminology [3] state that humans traditionally follow personal schedules within their own awareness spaces (e.g., favorite places). Modeling these routine activities may allow for new knowledge to be mined from movement datasets, potentially highlighting anomalous activities such as the visiting of new places (anomalous places) or the visiting of known places at uncommon times (anomalous visits). Not all patterns fitting this definition will be cause for concern, but highlighting them can be useful for interception and surveillance. We propose using the concept of return periods to quantify and identify anomalous places and anomalous visits in a targets movement data.

Figure 6 illustrates real-world GPS data of an individual over the course of 30 days. As described by routine activity theory, various routine locations are apparent which we labeled in the figure as “Work”, “Home”, and “Club”. The location labeled “Farm” was only visited once *(a return period of 30 days)*, illustrating a non-routine or anomalous place. In addition to anomalous places, we want to find anomalous visits. Table 2 displays “Arrival Count *(Return Period)*” for each of the four labeled places, distributed across four times of the day. For example, Cell (Work, Night) shows the target was present at Work at Night one time *(return period = 30 days)*. This is a deviation from routine activity, yet is in a known place. We label this pattern anomalous visits.

Modeling return periods for individuals in movement datasets, along with identifying anomalous behavior presents a number of challenges. Space and time partitioning of movement datasets
are needed to differentiate locations and time-slots of activity, along with the interaction (spatial and temporal auto-correlation) between these places. Statistical measures are needed to identify anomalous places and visits with confidence, in these GPS datasets. In addition, movement datasets can consist of millions and billions of GPS records, requiring scalable algorithms to efficiently find anomalous patterns.

Task R1: Develop a Conceptual Model for Denial-and-Deception Aware Return Periods (DDRP): Traditional return periods are associated with well-documented target phenomena (e.g., earthquakes, river floods). New models are needed for data-incomplete targets such as individuals practicing denial and deception. For example, the nightly total in Table 2 is 20 even though we expect 30 sightings since the dataset spans 30 days. We will investigate the following questions: What are taxonomies of denial beyond accidental and intentional? What are appropriate explicit representations of denied data and movement hypotheses generation (Section 5.3) in return periods for spatio-temporal events (e.g., (Home, Night))? 

Task R2: Investigate Interest Measures for Rare and Frequent Places and Visits: Robust interest measures are needed to accurately determine return periods of locations given incomplete information. This will require accounting for hypothesized movement (see previous section) in return period estimations. We will explore interest measures for finding anomalous places and anomalous visits in movement datasets. These measures should provide statistical confidence to assist in human interpretability and usefulness. They should also account for spatial and temporal auto-correlation as the dataset will be artificially partitioned.

Task R3: Develop Scalable Algorithms for Spatial Big Data: We will develop scalable algorithms for handling movement data in Spatial Big Data. These algorithms will need to efficiently utilize the computational structure of spatio-temporal auto-correlation to remain scalable will adding these traditionally expensive computations. We intend to explore parallel methods (MapReduce, GPU) methods to quickly process extremely large GPS datasets.

Task R4: Validation with Case-Study: The proposed algorithms and data structures will be evaluated using theoretical and experimental methodologies. A real world dataset supplied by BI, Inc. will allow us to validate the proposed algorithms’ usefulness to domain experts. In addition, we will use synthetic dataset tests to ensure scalability for Spatial Big Data.

6 Management

Project success will be measured in terms of (i) successful G.I.Sc. research resulting in the creation of models for identifying evasive patterns, (ii) the building of tools embodying the new results, and their adoption by GEOINT analysts, (iii) the success in being able to analyze and
Table 3: Project Task Schedule

<table>
<thead>
<tr>
<th>Quarters</th>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
</tr>
<tr>
<td><strong>Scientific Approach</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space-Time Entropy Discrimination</td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>Blackhole and Patterns of Evasion Detection</td>
<td>B1</td>
<td>B2</td>
</tr>
<tr>
<td>Theory-based Movement Hypothesis Generation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Return Period in Movement Datasets</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Management Approach</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress Monitoring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGA Reports</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Report</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

predict target movements and schedules.

In the third, fourth, and fifth optional years, we will explore the tools and the test bed to handle new blackhole patterns, more advanced movement hypotheses, and temporally-correlated return periods. Additional research tasks and work-packages will be determined in consultation with NGA NURI program managers and GEO INT Analysts.

The PI, Dr. Shekhar, will provide overall leadership for the project. He has experience in leading similar and larger size research teams. Graduate and undergraduate students will be working on the research and development of the methodologies proposed in the project.

7 Facilities and Equipment

The PI has access to outstanding computing facilities and resources within the University of Minnesota’s Department of Computer Science and Engineering. These include two SMP clusters each of which has with 16 Intel Itanium-2 CPUs running at 1 GHz and 32 GB of memory, 66 Dell PCs and 32 Macs (More details are available from http://www.lcse.umn.edu/). In addition, the Minnesota Supercomputing Institute has an IBM BladeCenter Cluster, SGI Altix, IBM Power4, IBM Netfinity Linux Cluster, and a Unisys ES7000 Orion 430 (More details are available at http://www.msi.umn.edu/). In addition, we have multiple relevant datasets for the proposed work from a number of collaborators (see the Collaboration section).

8 Sub-awards and Collaboration

Mark Abrams, a consultant for the National Reconnaissance Office with experience in GPS-deprived areas, and Prof. May Yuan at the University of Oklahoma working on movement pattern analysis, have offered their domain expertise. In addition, the University of Minnesota has an agreement with BI Incorporated, the makers of ankle-bracket GPS tracking devices for felons out on parole, to utilize their anonymized GPS-track datasets of offenders for research purposes.

9 Research Related Opportunities and Institutional Improvement

This grant will enhance capacities of University of Minnesota in the G. I. Sc. areas. A direct outcome of this research project will be the training and development of G. I. Sc. graduate students. Two Ph.D. students will be supported by research assistantships. The research will provide G.I.Sc.
Table 4: Summary of Research Team and Planned Commitments

<table>
<thead>
<tr>
<th>Professional Commitments</th>
<th>Research Team (% full-time)</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Proposal</td>
<td></td>
</tr>
<tr>
<td><strong>PI:</strong> Identifying and Analyzing Patterns of Evasion</td>
<td>9%</td>
</tr>
<tr>
<td>Sponsored Projects</td>
<td></td>
</tr>
<tr>
<td><strong>Co-PI:</strong> Expedition: Understanding Climate Change: A Data Driven Approach</td>
<td>9%</td>
</tr>
<tr>
<td><strong>Co-PI:</strong> Datanet: Terra Populus: A Global Population Environment Data Network</td>
<td>2.25%</td>
</tr>
<tr>
<td><strong>PI:</strong> Dynamic Purpose Aware graph models for composite networks (<strong>Ends 2014</strong>)</td>
<td>9%</td>
</tr>
<tr>
<td><strong>PI:</strong> Spatio-Temporal Pattern Mining for Multi-Jurisdiction Multi-Temporal Activity Datasets (<strong>Ends 2013</strong>)</td>
<td>9%</td>
</tr>
<tr>
<td>Other Work Commitments</td>
<td></td>
</tr>
<tr>
<td>University related teaching and professional services</td>
<td>Balance</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

projects for the Undergraduate Research Opportunity Programs and undergraduate honors theses. The research results will provide two weeks’ worth of materials in G.I.Sc. courses. Students will be encouraged to participate at relevant G.I.Sc. Conferences such as the annual summer assembly of the University Consortium on Geographic Information Sciences.

10 Current and Pending Support

The research proposed in this proposal has not been and will not be submitted to any other sources of funding during the evaluation period.

11 Research Team Description

The team at the University of Minnesota is experienced in spatial data mining with numerous publications [12]. Table 4 lists the names of all persons for whom financial support is proposed, the planned commitments (in units of percentage of full-time work year) to the proposed research, and the planned commitments to other work and professional activities.

The PI, Dr. Shekhar, will provide overall leadership for the project. He has experience in leading previous NGA grants with similar and larger-sized research teams. Dr. Shekhar actively publishes research papers, surveys and textbooks in the fields of spatial and spatio-temporal databases and data mining [9, 11, 12].
References

Abbreviated Curriculum Vitae for Shashi Shekhar

Shashi Shekhar is currently a Distinguished McKnight University Professor of Computer Science at the University of Minnesota, Minneapolis, MN, USA. He received the IEEE-CS Technical Achievement Award (2006) and was elected an IEEE fellow (2003) as well as an AAAS Fellow (2008) for contributions to spatial database storage methods, data mining, and geographic information systems (GIS). He co-edited an Encyclopedia of GIS (Springer, 2008, isbn 978-0-387-30858-6), and co-authored a textbook on Spatial Databases (Prentice Hall, 2003, isbn 0-13-017480-7) as well as over 250 research papers in peer-reviewed journals, books, conferences, and workshops. He is serving as a co-Editor-in-Chief of Geo-Informatica: An International Journal on Advances in Computer Sc. for GIS, a program co-chair for international conference on geographic information science (2012), a member of the National Research Council (NRC) GEOINT Workforce committee (2011) and a member of the Computing Community Consortium Council (2012-2015). He served as a general co-chair for the Symposium on Spatial and Temporal Databases (2011), and on the NRC Mapping Sciences Committee (2003-2009) of the National Academy, the Board of Directors of University Consortium of UCGIS (2003-2004), the editorial boards of IEEE Trans. on Knowledge and Data Engineering and the IEEE-CS Computer Science & Eng. Practice Board. Recent research accomplishments include co-location patterns for mining spatial databases, and scalable routing algorithms for evacuation planning. Earlier his group developed, CCAM, one of the most efficient storage methods for large road maps and scalable algorithms for computing shortest paths. More details at http://www.cs.umn.edu/~shekhar.

Affiliation: University of Minnesota, Dept. of Computer Science and Eng.

Mailing Address: 200 Union ST SE, #4-192, Minneapolis, MN 55455

Email: shekhar@cs.umn.edu URL: http://www.cs.umn.edu/~shekhar

Telephone: 612-624-8307 Fax: 612-625-0572

Professional Preparation

<table>
<thead>
<tr>
<th>Year</th>
<th>Degree</th>
<th>Field</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>Ph.D.</td>
<td>Computer Science</td>
<td>University of California, Berkley</td>
</tr>
<tr>
<td>1989</td>
<td>M.S.</td>
<td>Business Administration</td>
<td>University of California, Berkeley</td>
</tr>
<tr>
<td>1987</td>
<td>M.S.</td>
<td>Computer Science</td>
<td>University of California, Berkley</td>
</tr>
<tr>
<td>1985</td>
<td>B.S.</td>
<td>Computer Science</td>
<td>Indian Inst. of Tech., Kanpur, India</td>
</tr>
</tbody>
</table>

Appointments

<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005 -</td>
<td>Distinguished Univ. Professor</td>
<td>Univ. of Minnesota, Minneapolis, MN</td>
</tr>
<tr>
<td>2001 -</td>
<td>Professor</td>
<td>University of Minnesota, Minneapolis, MN</td>
</tr>
<tr>
<td>1995-2000</td>
<td>Assoc. Professor</td>
<td>University of Minnesota, Minneapolis, MN</td>
</tr>
<tr>
<td>1989-1995</td>
<td>Asst. Professor</td>
<td>University of Minnesota, Minneapolis, MN</td>
</tr>
</tbody>
</table>

Research Interests:

Data and knowledge engineering, spatial database management, spatial data mining, and geographic information systems.

Five Related Publications


Five Other Publications


Synergistic Activities

- Curriculum Development: Developed one of the first courses on Spatial Databases; Co-authored a popular textbook on Spatial Databases (Prentice Hall, 2003); co-edited an Encyclopedia of GIS (Springer, 2008), which was recommended highly by a review in ACM Computing Reviews (Nov. 2008); Presented tutorials on spatial data mining in conferences and other meetings; Lead a NSF IGERT on interdisciplinary graduate education around spatio-temporal issues (e.g., non-equilibrium dynamics); Chaired curriculum committee of Computer Science & Engineering department at the University of Minnesota (1998-2000); Served as a Computer Science representative on UCGIS curriculum committee (1998-99); Served on IEEE-Computer Society Computer Sc. and Eng. Practices Publication Board (1995-97).

- Active participation in broadening the participation of groups underrepresented in science via supervising over two dozen undergraduate (UG) students from historically black colleges in Army High Performance Computing Research Center annual summer workshops (1997-2006), NSF Research Experience for UGs, and UG Research Opportunity Program (UROP).