1. Program Name:
   “2007 NGA University Research Initiatives (NURI)”

2. BAA Number:
   “HM1582-07-BAA-0003”

3. Title:
   Spatio-Temporal Pattern Mining for Multi-Jurisdiction Multi-Temporal Activity Datasets

4. Research Topic:
   Topic 6: Analytic Tools and Techniques for Geospatial Information Science

5. 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.
   Andrea J. Martin, Grants Management Coordinator
   Phone:(612) 625-2480, Fax:(612) 625-0572, Email: amartin@cs.umn.edu

6. Signature of Principal Investigator. ___________________________ Date____________

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

8. Administrative Officer:
   Kevin McKoskey, Senior Grant Manager Sponsored Projects Administration,
   University of Minnesota, 200 Oak St SE, Suite 450, Minneapolis, MN 55455-2070,
   Phone: 612-624-5066; Fax: 612-624-5599, Email: awards@umn.edu

   Alternate contact information.
   Sowolo Abdullai, Grant Administrator, Sponsored Projects Administration,
   University of Minnesota, 200 Oak St SE, Suite 450, Minneapolis, MN 55455-2070,
   Phone: 612-626-2724; Fax: 612-624-5599, Email: abdu0092@umn.edu

9. Signature of Administrative Officer. ___________________________ Date____________
Space and time are crucial dimensions of human activities, such as crime. Thus, geospatial information sciences (G.I.Sc.), e.g. environmental criminology, have emerged in recent decades to develop understanding of the spatial distributions (e.g., high activity places or hot-spots) of activities as well as location-based factors affecting activities using analytical techniques, e.g. crime mapping, and analytical tools such as CrimeStat. However, in the last decade, the growth in variety and volume of observational data, notably multi-jurisdiction multi-temporal (MJMT) data, has out-paced the capabilities of analytical tools and techniques, and the ability of traditional human analysts to track, monitor, and predict human activities.

Major limitations of existing environmental criminology and spatial data mining models and techniques include the following. First, these do not adequately model richer temporal semantics of human activities (e.g. emerging hot spots, crime displacement). Second, these do not sufficiently explain the network location of activity hot-spots in terms of network properties. Third, these do not satisfactorily model heterogeneity across activity datasets from multi-modal transportation networks. These limitations represent critical barriers to progress of G.I.Sc., e.g. environmental criminology, since a significant fraction of human activity, particularly in cities as well as post-9/11 asymmetric urban warfare, revolves around transportation networks.

Proposed novel approaches to overcome these critical limitations include: i) providing a framework to model the rich semantics of the spatio-temporal (ST) patterns of human activities (e.g. emerging hot spots) by developing interest measures and scalable algorithms, ii) developing explanatory models of spatial network-based human activity patterns, and iii) developing an ST network ontology to integrate MJMT activity datasets from multi-modal transportation networks.

The proposed research has the potential to make a significant contribution in the area of research topic 6, namely, analytical tools and techniques for G.I.Sc. by advancing methods for tracking, monitoring, and predicting ST human activities in urban areas. Although the focus will be on analyzing crime, the resulting analytical tools and techniques will likely benefit the tracking, monitoring and predicting of many other kinds of ST human activities.

Our team, consisting of geographic information scientists and environmental criminologists, is capable of carrying out the proposed tasks. They not only have strong track records in G.I.Sc., data management, and human activity (e.g. crime) analysis but they have also worked collaboratively for the past two years. The PI, Dr. Shashi Shekhar, is a leader in ST data management and analysis. He will collaborate with Dr. Bhavani M. Thuraisingham and Dr. Latifur R. Khan who bring expertise on geospatial data analysis, secure heterogeneous data integration, and ontology-based models for information selection.

Collaborator Dr. Ned Levine is well-known for CrimeStat, a leading statistical software program for crime analysis. Technical Advisor Dr. Brian J.L. Berry, Professor and Dean of the School of Economics from the University of Texas, Dallas is the most cited geographer for more than 25 years. Collaborator Mille DeAnda serves as President of the North Texas Crime Commission and has 7 years work experience with law enforcement agencies, including the FBI. Dr. Budhendra Bhaduri is the Research Leader of the Geographic Information Science and Technology group at the Oak Ridge National Laboratory. Steven Seida, a principal engineer at Raytheon and Bari A. Lee, a Senior Intelligence Analyst at Fusion System can assist us in the technical research. The researchers and collaborators make this team truly unique.
BAA-4.2.3. RESEARCH PLAN

1. OBJECTIVES

The overall objective of this proposal is to model and mine spatio-temporal (ST) patterns in multi-jurisdiction and multi-temporal (MJMT) datasets to accurately track, monitor, and predict human activities. Several scientific and technical challenges arise when modeling ST patterns in MJMT datasets due to the spatio-temporal nature and heterogeneity of datasets. These represent major barriers to progress in the Geospatial Information Science (G.I.Sc.) fields such as environmental criminology. Thus, we propose to: (1) investigate semantically diverse ST activity patterns that frequently occur in MJMT datasets, (2) probe ST network patterns in urban datasets, and (3) explore common ST networks ontologies to effectively integrate MJMT datasets about urban activities centered around multi-modal transportation networks.

Current state of knowledge may be categorized into two main areas: environmental criminology [a,f] and spatial data analysis [b,c,d,e]. In environmental criminology, there are several ST theories such as Routine Activity Theory (RAT) and Crime Pattern Theory (CPT). RAT suggests that the location of a crime is related to the criminal’s frequently visited areas. CPT extends this theory on a spatial model shown in Figure 1. This spatial model consists of nodes (e.g. frequently visited areas such as home, work, entertainment/recreation), paths (e.g. routes between nodes), and edges (i.e. boundaries of an activity footprint). CPT suggests that crime locations are often close to the edges, i.e. near the criminal’s activity boundaries, where the residents may not recognize the criminal. Spatial data analysis techniques such as spatial statistics [b,c] and spatial data mining [d,e] have explored spatial regression, space-time interaction (e.g. Knox test), spatial clustering for hot spot detection, spatial outliers, co-location of subsets of crime-types, etc. Current methodologies have three main limitations: (1) they do not adequately model richer temporal semantics beyond space-time interaction, (2) they do not satisfactorily explain the cause of detected hot spot locations on spatial networks, such roads, trains, and (3) they do not effectively model heterogeneities across spatial networks, e.g. multi-modal urban transportation modes (such as light-rail subways and roads).

Thus, we propose novel approaches as first steps to address these limitations by (1) quantifying ST patterns via novel interest measures which yield computationally efficient pattern-mining algorithms, (2) extending explanatory models, e.g. spatial regression, to ST network spaces, and (3) developing ST network ontologies to integrate activity datasets across multi-modal transportation networks.

References

2. ANTICIPATED RESULTS

Anticipated results of this proposal include mathematical models, computer algorithms, data analysis methods, and analytical tools, e.g. new software or extensions to popular software, such as CrimeStat. Specifically, we anticipate the following results:

- **Mathematical models** to define interest measures to quantify sub-classes of ST activity patterns (e.g. periodic or emerging hot spots); and to design novel explanatory models, e.g. network-based spatial regression, for multi-modal ST network patterns.
- **Computer algorithms** to design new scalable ST activity pattern mining procedures. We also propose to develop algorithms to automate routine tasks of MJMT data integration.
- **Data analysis methods** to create a new classification of ST activity patterns (e.g. emerging hot spots, periodic hot spots, and moving hot spots, which may represent the concept of displacement in environmental criminology). We also propose to develop ontology of environmental criminology concepts for at least for a few selected crime types that utilize MJMT data integration process.
- **Analytical tools** to develop new software and extend the capabilities of the state-of-the-art software, e.g. CrimeStat, by including newly designed ST activity pattern mining algorithms, scalable network analysis algorithms, and ST network data integration methods.

3. APPLICABILITY

The critical barrier of modeling the MJMT data heterogeneities and rich pattern semantics limits environmental criminologists (e.g. many at the Ninth Crime Mapping Research Conference, 3/2007, http://www.ojp.usdoj.gov/nij/maps/pittsburgh2007/index.html) from quickly identifying many ST patterns, which are crucial for timely intervention for crime prevention. There are three significant limitations of current environmental criminology and spatial data analysis techniques creating this critical barrier. Let us look at these in further detail.

![Figure 2: Activity Levels by Jurisdiction](http://www.diligencellc.com) (Source: http://www.diligencellc.com) (Best Viewed in Color)
First, traditional approaches do not explicitly model temporal semantics such as trends or periodic patterns. For example, Figure 2, particularly the numeric activity count data, shows a diminishing trend for the number of insurgent incidents across multiple provinces from July 19-26, 2004 (Figure 2a) and July 26-Aug 2, 2004 (Figure 2b). Notice the highlighted entries in numeric activity count data in Figure 2, e.g. Anbar, where the number of insurgent incidents diminished in a matter of weeks. Likewise, Figure 3 gives an example of a periodic activity pattern, where a high level of activity seems to periodically occur between 2300 and 0000 hours each day of a week. Timely identification of such ST patterns is crucial for improving public safety. However, it takes enormous amount of time and human effort to identify ST patterns using current tools and techniques, particularly for MJMT datasets. Thus, we propose novel ST pattern mining methods in Section 4.1.1 and MJMT data integration methods in Section 4.1.3.

The second major limitation is inadequacy of explanatory models for activity around ST networks such as train-networks, road-networks, etc. For example, Figure 4 illustrates the incidents along major roads from July 26-Aug 2, 2004 (Figure 4a) and Aug 2-9, 2004 (Figure 4b). Public safety professionals may be interested in analyzing the ST network factors to explain high activity levels or changes in activity levels at certain highway segments to compare prevention options, e.g. check-points. Such analysis is not only arduous using current methodologies, but may not be statistically meaningful since common methods, e.g. spatial regression, do not adequately model ST network constraints, such as connectivity, one-ways streets, etc. Thus, we propose to explore a novel ST network analysis method to study explanatory models for ST network patterns in Section 4.1.2.

Another significant limitation relates to the inadequate ability to quickly integrate and utilize activity data reported across multiple jurisdictions or timeframes. Figure 4 shows activity levels on segments of a road network. In addition, there may be other transportation-mode, e.g. train, based datasets, which may be useful for understanding global activity patterns. Such analysis
over combined activity datasets is very laborious today due to MJMT data heterogeneities. Thus, we propose to explore common ST network ontologies to unify activity data from multiple modes of transportation networks when discovering novel ST network patterns (Section 4.1.3). Another example of data heterogeneity occurs in Figure 2, since color-codes used in Figures 2(a) and 2(b) are not directly comparable. In fact, a casual observer may erroneously infer an increasing trend in activity level in the Anbar province over time using the color codes, if they did not notice the numeric data and definitions of color codes.

Our proposed approaches also address several important issues listed under “Desired Research” in Topic 6 of BAA. First, “...integrating and analyzing heterogeneous geospatial data sources...”1 analyzed independently in criminal analyses is addressed in section 4.1.3. Second, “…the probability of a criminal activity occurring at a particular location, given its prior occurrence...” and “…account for phenomena such as time delays, periodic or permanent movement of variables...” is studied using spatial hot spots by environmental criminology which limits the discovery of temporal patterns across space and is addressed in section 4.1.1. Third, “…data pertaining to one variable be used to draw conclusions about a different variable...” is examined by using spatial regression in environmental criminology to explain hot spots via location-based and ST network-based factors and is addressed in section 4.1.2.

4. PROPOSED APPROACH TO ACHIEVE OUR OBJECTIVES

This section presents scientific (Section 4.1) and management (Section 4.2) approaches.

4.1. SCIENTIFIC APPROACH

The scientific and technical challenges outlined in Section 3 are addressed in section 4.1. Section 4.1.1 presents proposed ST activity pattern mining approach and its tasks, section 4.1.2 describes our ST network pattern mining approach and its tasks, and section 4.1.3 describes multi-jurisdiction multi-temporal (MJMT) data integration and its tasks.

4.1.1. Spatio-temporal Activity Pattern Mining

ST activity pattern mining is an important issue in the domain of environmental criminology, since location and time are important dimensions of human activity. Relevant locations include criminal’s residence, incident location, criminal’s workplace, etc. Several types of ST activity patterns are explored in environmental criminology. Some of these pattern-types are trends (e.g. reduction of activity level in Anbar in Figure 2) and periodic activity patterns (e.g. periodic high-activity level between 2300 and 0000 in Figure 3).

Current state-of-the-art techniques in environmental criminology have focused on discovering spatial activity patterns (e.g. hot spots, spatial regression), or space-time interaction (e.g. Knox test). However, they do not adequately model rich temporal semantics such as trends, periodicity, etc. Thus, we propose to model ST activity patterns and design computationally efficient novel pattern mining methods.

We plan to start the ST pattern discovery process with the categorization of diverse temporal semantics of activity patterns. Then, we quantify selected pattern-categories via interest measures and design computationally efficient algorithms to quickly mine instances of these

1 Italic quotations denote direct quotes from the RFP in Task 6 (1.2.6).
pattern-categories from massive MJMT datasets. We plan to evaluate the proposed algorithms with real and synthetic datasets and compare them with traditional data analysis methods. Finally we explore ways of generalizing the proposed pattern-categories, interest measures, and algorithms to other application domains such as public health. The proposed research tasks are summarized as follows:

**Task T1. Classification of semantically diverse ST activity patterns:** We propose to create a classification of ST activity patterns in collaboration with environmental criminologists. The new sub-types of ST activity patterns may include: (1) *Periodic activity patterns* such as the daily high activity level between 2300 and 0000 hours in Figure 3; (2) *Moving hot-spot patterns* such as crime displacement to new locations due to increased law enforcement effort at a location; (3) *Diminishing activity patterns* as in the province Anbar in Figure 2; (4) *Spatial-time series correlation pattern*, e.g. weather-events may affect activity levels; and (5) *Ordered co-incidence activity patterns* where a certain sequence of events leading to a criminal event may be of interest.

**Task T2. Quantitative interest measures:** We plan to explore and evaluate alternative interest measures to quantify major categories of ST activity patterns. Ideal interest measures should not only characterize a category of ST patterns, but also lead to computationally efficient algorithms based on properties like anti-monotone.

**Task T3. Design data analysis algorithms for timely mining of ST activity patterns:** We will design new scalable ST activity pattern mining algorithms, via quick pruning of irrelevant candidates using properties of quantitative interest measures. These algorithms should not only be computationally efficient, but also yield correct and complete results.

**Task T4. Self-Evaluation:** 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 computational efficiency of proposed algorithms under different combinations of parameters, such as dataset size, number of activity patterns, and thresholds on interest measures.

**Task T5. Comparative-Evaluation:** We plan to compare proposed algorithms with traditional spatial data analysis methods (e.g. Knox-test, Markov model) on common datasets. A key goal is to test if proposed interest measures and algorithms can identify novel, useful and interesting ST patterns, which are not quickly identified by traditional methods.

**Task T6. Generalize to activities other than crime:** We will evaluate proposed ST pattern categories, interest measures and algorithms for effectiveness in other application domains such as to public health and water quality.

### 4.1.2. Spatio-temporal Network Pattern Mining

In post-9/11 world of asymmetric warfare in urban area, many human activities are centered about ST infrastructure networks, such as transportation, oil/gas-pipelines, and utilities (e.g. water, electricity, telephone). Thus, activity reports, e.g. crime/insurgency reports, may often use network based location references, e.g. street address such as “200, Quiet Street, Scaryville, RQ 91101”. In addition, spatial interaction among activities at nearby locations may be constrained by network connectivity and network distances (e.g. shortest path along roads or train networks) rather than geometric distances (e.g. Euclidean or Manhattan distances) used in traditional spatial analysis. Crime prevention may focus on identifying subsets of ST networks with high activity
levels, understanding underlying causes in terms of ST-network properties, and design ST-
network-control policies.

Since discovery of ST network patterns in MJMT datasets can play a key role in urban areas,
environmental criminology has started exploring the *Journey-to-Crime* analysis to identify
frequently traveled public-transportation routes by criminals. Figure 5 illustrates the *Journey-to-
Crime* pattern for Baltimore, MD for activities related to local-train network using the CrimeStat
software. Figure 5a shows the available train routes and Figure 5b shows the spatial locations of
the criminals’ residences (based on arrest records) and locations of crimes. It also shows the
possible <origin, destination> pairs for each crime assigned to the train network. Figure 5c gives
an example output of the *Journey-to-Crime* estimates from the CrimeStat software. Thick lines in
Figure 5(c) show those train-network segments, which are used more often by criminals to reach
crime locations than other segments.

![Figure 5: Journey-to-Crime Estimates via the CrimeStat software (Best Viewed in Color)](image)

As mentioned in Section 3, existing spatial analysis methods face several challenges. First,
these methods do not model the effect of explanatory variables to determine the locations of
network hot spots. Second, existing methods for network pattern analysis are computationally
expensive. For example, *Journey-to-Crime* patterns are often determined based on the shortest
paths from the criminals’ residences to the crime locations. Shortest path computation could be
the most expensive step in the discovery process, especially in cases where the size of the
network and the number of crime locations of interest are substantially large. Third, these
methods do not consider the temporal aspects of the activity in the discovery of network patterns.
For example, the routes used by criminals during the day and night may be different. The
periodicity of the bus/train schedules can have an impact on the routes traveled. Incorporating
the time-dependency of the transportation network can improve the accuracy of the patterns.

We plan to investigate explanatory models for ST Network activity patterns to address the
following issues. In traditional spatial regression models, neighborhood is defined by the spatial
proximity of locations. For spatial network patterns this may not be adequate since network-
distance (e.g. shortest path among nodes) may be quite different than straight-line or Manhattan
distance. Thus, we propose to probe generalization of spatial regression models to describe the
effect of explanatory variables on spatial network patterns.

Another foci in the proposed research would be to design computationally efficient
algorithms for *journey-to-crime* analysis and to incorporate the temporal aspects in the discovery
of network patterns. Inclusion of temporal aspects in a discovery process leads to some unique
challenges. Tracking the time-dependence of the network parameters and criminal behavior can
result in a large volume of data. This suggests the need for a storage efficient data model that can
represent the ST nature of the networks. Path computations on time-dependent networks operate on large search spaces and effective heuristics might be crucial in developing computationally efficient and scalable algorithms. We propose to design computationally efficient, scalable, and explanatory algorithms to mine these patterns from massive MJMT datasets. We evaluate the proposed algorithms with real and synthetic datasets and compare them with existing algorithms. Finally we will explore the possibility of generalizing the proposed algorithms for different domains. The proposed research tasks are summarized as follows:

**Task S1. Network-based Spatial Regression:** We plan to investigate alternative definitions of neighborhood of spatial network variables. We also plan to develop scalable parameter-estimation methods spatial regression techniques to explain spatial network patterns.

**Task S2. Scalable methods for Journey-to-Crime Patterns:** We plan to probe ways to improve the scalability of network algorithms for journey-to-crime analysis. In prior work, we explored various shortest path algorithms in conjunction with several data structures in the context of static, road networks. We propose to extend this work for the temporal aspects and transfer result to CrimeStat software to speed up its “Journey to Crime” functionality.

**Task S3. ST models for Network Systems:** We plan to develop a model that represents time dependence in transportation networks. In addition, we will extend these models to include multi-modal transportation networks. This model would be used to design algorithms for the discovery of interesting patterns.

**Task S4. Generalization of ST Network Patterns:** We plan to explore generalization of results from section 4.1.1 to ST networks to explore network patterns such as trends (e.g. displacement of activity hot-spots within a route or across intersecting routes), network based outliers (e.g. bus-stops which are different from upstream and down-stream bus-stops) etc. We propose to develop ST network based interest measures and scalable algorithms to find these patterns.

**Task S5. Validation and Verification:** We plan to evaluate our proposed approaches through comparisons of existing approaches, via analytical and experimental evaluation in collaboration with application-domain scientists.

4.1.3. Spatio-temporal Data Integration

Many environmental criminology techniques assume that data are locally maintained and the dataset is homogeneous as well as certain. This assumption is not realistic as ST data is often managed by different jurisdictions and therefore, the analyst may have to spend unusually large amount of time to link related events across different jurisdictions (e.g., the sniper shootings across Washington DC, Virginia and Maryland in October 2002).

Two major challenges that need to be addressed when integrating heterogeneous crime data sources are semantic heterogeneity, and data uncertainty. Semantic heterogeneity occurs when there is a definition-mismatch across MJMT datasets. Consider the multi-modal transportation networks illustrated in Figure 6. Even though each network may be modeled using graph theory, the definition of “nodes” and “edges” may differ across transportation modes. For example, nodes may represent road intersections in a road network but “stops” in bus/train lines in other networks. In addition, the location-specifications may differ across transportation modes. For example, bus routes may use street addresses in urban areas, or (highway, mile-marker) on major roadways. However, train networks may specify location by train-routes and stops. Conflating networks across transportation modes for multi-modal analysis is a demanding task. For
example, semantic differences across transportation modes make it difficult for web-based routing services (e.g., mapquest.com) to offer multi-modal routes involving walking, driving, trains, and buses, since current data-integration approaches (e.g., GML, wrappers, ST ontologies) do not adequately model ST network concepts and heterogeneities. In summary, lack of integrated spatial network based activity datasets will make it tedious to perform analysis for MJMT activity patterns. Thus, we propose the development of ST network ontologies as a first step toward meeting the challenges in integrating heterogeneous data sources. These ST network ontologies will be integrated with other ontologies that provide, for example, definitions of various environmental criminology terms.

The introduction of uncertainty into multi-modal ST network systems creates additional challenges for the data integration problem. Sources of uncertainty may include motion or travel on a public transportation vehicle. In addition, the activity-location uncertainties may differ across transportation modes due to variation in map accuracy standards, geo-coding algorithms, and sampling rates of automatic vehicle location sensors such as GPS. Thus, we propose to investigate methods to handle uncertainty across multi-modal transportation network systems.

We will first develop common ontologies across multiple ST networks towards a data integration framework. This is followed by an uncertainty management system to handle diverse location-uncertainties across multi-modal transportation systems. Proposed ideas will be evaluated using real datasets. The proposed research tasks can be summarized as follows:

**Task H1. Design ontology for ST Network activities:** We plan to investigate common ontologies for multi-modal transportation networks to integrate multiple sources for network data.

**Task H2. Develop a data integration framework for ST network datasets:** We will extend our research on the geospatial semantic web to develop a data integration framework for ST network datasets. We will utilize the service oriented architecture framework upon which the Global Information Grid is based so that our framework can be integrated with frameworks being developed for network centric enterprise services.

**Task H3. Data Integration algorithms and tools:** We will probe algorithms to automate routine tasks in MJMT data integration leveraging the common ST-Network ontology.

**Task H4. Data Uncertainty Management for Multi-Modal Transportation Systems:** We will investigate models to represent diverse uncertainty representations across ST network datasets. These techniques will subsequently be integrated into our data integration framework and tools.

**Task H5. Validation and Verification:** We plan to evaluate our proposed approaches through the use of real world datasets from multiple jurisdictions and transportation modes.
4.2 MANAGEMENT APPROACH

We will measure the success of this project in terms of (i) successful GI.Sc. research resulting in the creation of new ST data analysis techniques, (ii) the building of tools embodying the new results, and their use by analysts of environmental criminologists, (iii) the success in being able to address MJMT human activity data analysis questions that were previously intractable.

The main milestones at the end of the first year will include: (1) the development of a data integration test bed to handle the semantic heterogeneity of ST-network datasets; (2) a toolkit consisting of semantically diverse ST pattern mining methods; and (3) a ST network pattern mining toolkit. At the end of the second year, the main milestones will include: (1) a prototype system that will demonstrate the data analysis tools on the interoperable test bed using two jurisdictions; (2) novel approaches for data analysis (e.g. ST activity and network patterns); and (3) the development of a model to handle uncertainty in MJMT datasets in order to incorporate uncertainty reasoning into the interoperability test bed. The breakdown of the research tasks is given in Table 1, which lists the milestones for quarterly intervals and project management activities to track the progress of this project. Milestones are categorized based on the three work-packages proposed in Sections 4.1.1, 4.1.2 and 4.1.3. Each work-package is subdivided into well-defined quarterly tasks specified in Section 4.1. The last quarter is reserved for preparation of the final report.

<table>
<thead>
<tr>
<th>Quarters</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Final Report</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td>Scientific Approach</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatio-temporal Activity Pattern Mining (4.1.1)</td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
</tr>
<tr>
<td>Spatio-temporal Network Pattern Mining (4.1.2)</td>
<td>S1</td>
<td>S2</td>
<td>S2</td>
</tr>
<tr>
<td>Spatio-temporal Data Integration (4.1.3)</td>
<td>H1</td>
<td>H1</td>
<td>H2</td>
</tr>
<tr>
<td>Management Approach</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team Collaboration</td>
<td>Bi-weekly conference calls, E-mail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress Monitoring</td>
<td>Quarterly review of goals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGA Reports</td>
<td>Six month progress report to NGA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Project Task Schedule (Tasks T1…T6, S1…S5, H1…H5 described in Section 4.1)

In the third, fourth, and fifth optional years, the tools and the test bed to handle multiple justifications and different types of datasets will be enhanced and utilized. Additional research tasks and work-packages will be determined in consultation with NGA NURI program managers and researchers from environmental criminology.

The PI, Dr. Shekhar, will provide overall leadership for the project. He has experience in leading similar and larger size research teams. The Co-PI, Dr. Thuraisingham, will provide leadership for the effort at U. T. Dallas. She has considerable leadership and technical experience in academia and industry (e.g. MITRE, Honeywell) and Government (NSF, CIA, NSA, AF, Navy, Army and the IRS). Dr. Shekhar will also guide the scientific research tasks described in sections 4.1.1 and 4.1.2. Dr. Thuraisingham and Dr. R. Khan will guide the scientific research specified in Section 4.1.3. Graduate and undergraduate students will be working on the research and development of the methodologies proposed in this project.
5. FACILITIES AND EQUIPMENT

The departments of computer science at the University of Minnesota and the University of Texas at Dallas have several workstations and servers available for the proposed research. However, we will need high-resolution monitors on systems equipped with sophisticated video cards and high memory cards for visualization of large and complex MJMT datasets and validation of our results. Real crime datasets provided by Dr. Ned Levine will be used as input data to test our techniques.

6. SUB-AWARDS AND COLLABORATION

We plan to have a sub-award to the University of Texas at Dallas with Dr. Thuraisingham and Dr. Khan. Both researchers will concentrate on the data integration portion of this proposal (section 4.1.3) and use our research results in their own research areas.

We plan to collaborate with a commercial firm, namely, Dr. Ned Levine and Associates. This will provide a means to disseminate research results. Specifically, the CrimeStat software may be a technology transfer vehicle for the anticipated results. In addition, he will contribute by providing current state-of-the-art crime analysis tools, domain specific feedback on proposed G.I.Sc. theories, and additional criminal activity datasets. Dr. Budhendra Bhaduri will provide guidance on data-integration tasks based on ORNL (Oakridge National Laboratory) experience with large data-integration projects in context of global population datasets. Dr. Brian Berry can provide guidance on Geography, spatial analysis and social science. Millie DeAnda can bring in her rich experience with the law enforcement agencies throughout the country and provide us domain knowledge about crime. Steven Sieda and Bari A. Lee can provide us feedback on the relevance and usefulness of results for Homeland Security domain.

7. STUDENT TRAINING AND INSTITUTIONAL IMPROVEMENT

This grant will enhance capacities of University of Minnesota and University of Texas (Dallas) 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. Through this project, they will interact with team members from diverse domains and learn valuable skills of communicating and contributing to G. I. Sc. research.

The project will also enhance undergraduate G.I.Sc. education. The research will provide G.I.Sc. 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 annual summer assembly of the University Consortium on Geographic Information Sciences.

8. 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.
Table 2 lists the names of all persons for whom financial support is proposed, the planned commitments (in units of a percentage of full-time work year) to the proposed research, and the planned commitments to other work and professional activities.

<table>
<thead>
<tr>
<th>Research Team</th>
<th>Planned Commitments (% full-time work year)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>University of Minnesota</td>
</tr>
<tr>
<td></td>
<td>Faculty</td>
</tr>
<tr>
<td>Projects</td>
<td></td>
</tr>
<tr>
<td>This Proposal</td>
<td>Shashi Shekhar</td>
</tr>
<tr>
<td>Spatio-Temporal Pattern Mining for Multi-Jurisdiction Multi-Temporal Activity Datasets, Submitted to NGA-NURI, 09/01/2007 - 08/31/2012</td>
<td>9%</td>
</tr>
<tr>
<td>1. Modeling and Mining Spatio-temporal Data, Army Corps of Engineers, 03/15/2006 - 09/30/2009, $112,000.</td>
<td>5%</td>
</tr>
<tr>
<td>3. Information Operations Across Infospheres, Air Force Office of Scientific Research, 01/2006 - 12/2008, $300,000.</td>
<td></td>
</tr>
<tr>
<td>4. Design and Development of Semantic Web and Data Mining Technologies for Geospatial Data, Raytheon, 06/2006 - 06/2007, $100,000.</td>
<td></td>
</tr>
<tr>
<td>Other Work and Professional Activities</td>
<td>University related teaching and professional services</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 2: Summary of Research Team and Planned Commitments