REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
12-08-2016	Final Report	29-Apr-2009 - 28-Apr-2015

4. TITLE AND SUBTITLE	
Final Report: PAGE: Policy Analytics Generation Engine	5a. CONTRACT NUMBER W911NF-09-1-0206
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS V.S. Subrahmanian	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Maryland - College Park Office of Research Administration 3112 Lee Building 7809 Regents Drive College Park, MD 20742 -5141	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO 11. SPONSOR/MONITOR'S REPORT NUMBER(S) 56057-NS.81

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

SOMA (Stochastic Opponent Modeling Agents) are a paradigm within which it is possible to make statements of the form "If condition C is true in the environment in which a group G operates, then group G will take action A with probability in the range [L,U]". SOME has been used to model the behaviors of various terrorist groups and make forecasts about future attacks by those groups. The PAGE project uses SOMA models of group behavior in order to generate policies that achieve a desired goal. A policy is a way of changing the environment in which the

15. SUBJECT TERMS

Stochastic Opponent Modeling Agent, SOMA, Geospatial Abduction, Betweeness Centrality, Covertness Centrality

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Venkatramanan Subrahmanian
UU	UU	UU	UU		19b. TELEPHONE NUMBER 301-405-6724

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report: PAGE: Policy Analytics Generation Engine

ABSTRACT

SOMA (Stochastic Opponent Modeling Agents) are a paradigm within which it is possible to make statements of the form "If condition C is true in the environment in which a group G operates, then group G will take action A with probability in the range [L,U]". SOME has been used to model the behaviors of various terrorist groups and make forecasts about future attacks by those groups. The PAGE project uses SOMA models of group behavior in order to generate policies that achieve a desired goal. A policy is a way of changing the environment in which the group operates (subject to feasibility constraints). We develop such techniques on sequential machines and then develop a parallel framework for it. We also developed policies and methods by which a group of defensive resources (e.g. checkpoints) could be situated in a given geography in order to minimize attacks by an adversary on both static and moving targets (e.g. convoys). In addition, the project looked at the problem of computing between centrality in hypergraphs – also called Group Between-Ness Centrality. GBC describes the probability that a node or a group of nodes lie on a randomly chosen shortest path between two randomly selected nodes. We developed algorithms to scalably compute both BC and GBC in huge networks, outperforming (under certain conditions) all previous results. We also proposed the concept of “covertness centrality” which identifies how a “bad guy” can try to hide from “good guys” in a social network without being easily identified – by understanding how such “bad guys” can hid in social networks, we have learned how to identify them.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

08/24/2012 34.00 Gerardo I. Simari, John P. Dickerson, Amy Sliva, V.S. Subrahmanian. Parallel Abductive Query Answering in Probabilistic Logic Programs,
ACM Transactions of Computational Logic, (01 2011): 0. doi:

08/24/2012 35.00 Matthias Broecheler, V.S. Subrahmanian, Cristian Molinaro, Paulo Shakarian. Using Generalized Annotated Programs to Solve Social Networks Diffusion Optimization Problems,
ACM Transactions of Computational Logic, (01 2011): 0. doi:

TOTAL: 2

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper
07/09/2013 50.00	Amos Azaria, Ariella Richardson, Sarit Kraus. Autonomous Agent for Deception Detection, AAMAS Conference 2013. 06-MAY-13, . : ,
07/09/2013 49.00	Noam Peled, Kobi Gal, Sarit Kraus. Learning to Reveal Information in Repeated Human-Computer Negotiation, Human-Agent Interaction Design and Models Workshop 2012. 04-JUN-12, . : ,
07/09/2013 51.00	Amos Azaria, Ariella Richardson, Avshalom Elmalech, Avi Rosenfeld, Sarit Kraus, David Sarne. On Automated Agents' Rationality, AAMAS Conference 2013. 06-MAY-13, . : ,
07/09/2013 52.00	Amos Azaria, Sarit Kraus. Advice Provision in Multiple Prospect Selection Problems, AAMAS Conference 2013. 06-MAY-13, . : ,
07/09/2013 56.00	Chanhyun Kang, Andrea Pugliese, John Grant, V.S. Subrahmanian. STUN: Querying Spatio-Temporal Uncertain (Social) Networks, ACM International Conference on Advances in Social Networks Analysis and Mining. 26-AUG-12, . : ,
07/09/2013 42.00	Samuel Barrett, Peter Stone, Sarit Kraus, Avi Rosenfeld. Teamwork with Limited Knowledge of Teammates, AAAI Conference 2013. 14-JUL-13, . : ,
07/09/2013 39.00	Noam Peled, Ya'akov Gal, Sarit Kraus. An Agent Design for Repeated Negotiation and Information Revelation with People, National Conference on Artificial Intelligence 2013. 14-JUL-13, . : ,
07/09/2013 40.00	Noam Hazon, Raz Lin, Sarit Kraus. How to Change a Group's Collective Decision?, International Joint Conference on Artificial Intelligence. 03-AUG-13, . : ,
07/09/2013 41.00	Noam Peled, Moshe Bitan, Joseph Keshet, Sarit Kraus. Predicting Human Strategic Decisions Using Facial Expressions, International Joint Conference on Artificial Intelligence 2013. 03-AUG-13, . : ,
07/09/2013 43.00	Moshe Bitan, Ya'akov Gal, Sarit Kraus, Elad Dokow, Amos Azaria. Social Rankings in Human-Computer Committees, AAAI Conference 2013. 14-JUL-13, . : ,
07/09/2013 44.00	Thanh Nguyen, Rong Yang, Amoms Azaria, Sarit Kraus, Milind Tambe. Analyzing the Effectiveness of Adversary Modeling in Security Games, AAAI Conference 2013. 14-JUL-13, . : ,
07/09/2013 45.00	Albert Xin Jiang, Zhengyu Yin, Chao Zhang, Milind Tambe, Sarit Kraus. Game-theoretic Randomization for Security Patrolling with Dynamic Execution Uncertainty, AAMAS Conference 2013. 06-MAY-13, . : ,
08/12/2016 68.00	. Covertness Centrality in Networks, FOSINT-SI. 27-AUG-12, Istanbul, Turkey. : ,

08/12/2016 69.00 . Abductive Inference for Combat: Using SCARE-S2 to Find High-Value Targets in Afghanistan, International Conf. on Innovative Applications of Artificial Intelligence 2011. 09-AUG-11, San Francisco, CA. : ,

08/12/2016 71.00 . Abductive Inference in Probabilistic Logic Programs, Technical Communications of the Int'l Conference on Logic Programming 2010. 16-JUL-10, Edinburgh, Scottland. : ,

08/12/2016 73.00 . Approximate Achievability in Event Databases, The 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty. 29-JUN-11, Belfast, Northern Ireland, UK. : ,

08/12/2016 75.00 . Efficient Multi-View Maintenance in the Social Semantic Web, 21st World Wide Web Conference 2012. 16-APR-12, Lyon, France. : ,

08/12/2016 76.00 . STUN: Spatio-Temporal Uncertain (Social) Networks, ASONAM 2012. 26-AUG-12, Istanbul, Turkey. : ,

08/12/2016 78.00 . Diffusion Centrality in Social Networks, ASONAM 2012. 26-AUG-12, Istanbul, Turkey. : ,

TOTAL: **19**

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

03/09/2012 5.00 Austin Parker, V.S. Subrahmanian, John Grant. Fast and Accurate Prediction of the Destination of Moving Objects, SUM 2009. 30-SEP-09, . : ,

03/09/2012 27.00 Massimiliano Albanese, Cristian Molinaro, Fabio Persia, Antonio Picariello, V.S. Subrahmanian. Finding "Unexplained" Activities in Video, 22nd International Joint Conference on Artificial Intelligence . 22-JUN-11, . : ,

03/09/2012 20.00 Paulo Shakarian, Margo K. Nagel, Brittany E. Schuetzle, V.S. Subrahmanian. Abductive Inference for Ccombat: Using SCARE-S2 to Find High-Value Targets in Afghanistan, 23rd Innovative Applications of Artificial Intelligence Conference 2011. 08-AUG-11, . : ,

05/01/2012 13.00 Galit Haim, Ya'akov Gal, Sarit Kraus, Michele J. Gelfand. A Cultural Sensitive Agent for Human-Computer Negotiation, 11th Annual International Conference on Autonomous Agents and Multiagent Systems 2012. 08-JUN-12, . : ,

08/24/2012 37.00 Andrea Pugliese, John Grant, V.S. Subrahmanian, Chanhyun Kang. STUN: Spatio-Temporal Uncertain (Social) Networks, ACM/IEEE ASONAM (Analysis of Social Networks and Media). 26-AUG-12, . : ,

TOTAL: **5**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper
03/09/2012	1.00	Gerardo I. Simari, John P. Dickerson, Amy Sliva, V.S. Subrahmanian. Parallel Abductive Query Answering in Probabilistic Logic Programs, ACM Transactions of Computational Logic (01 2011)
03/09/2012	26.00	Massimiliano Albanese, Rama Chellapa, Naresh Cuntoor, Vincenzo Moscato, Antonio Picariello, V.S. Subrahmanian. PADS: A Probabilistic Activity Detection Framework for Video Data, IEEE Transactions on Pattern Analysis and Machine Intelligence (12 2010)
03/09/2012	24.00	Massimiliano Albanese, Andrea Pugliese, V.S. Subrahmanian. Fast Activity Detection: Indexing for Temporal Stochastic Automation based Activity Models, IEEE Transactions on Knowledge & Data Engineering (12 2011)
03/09/2012	23.00	Amos Azaria, Yonatan Aumann, Sarit Kraus. Automated Strategies for Determining Rewards for Human Work, Association for the Advancement of Artificial Intelligence (02 2012)
03/09/2012	22.00	Paulo Shakarian, Matthias Broecheler, V.S. Subrahmanian, Cristian Molinaro. Using Generalized Annotated Programs to Solve Social Network Diffusion Optimization Problems, ACM Transactions of Computational Logic (01 2011)
03/09/2012	19.00	Paulo Shakarian, V.S. Subrahmanian, Maria Luisa Sapino. GAPs: Geospatial Abduction Problems, ACM Transactions on Intelligent Systems and Technology (10 2011)
03/09/2012	17.00	Paulo Shakarian, John P. Dickerson, V.S. Subrahmanian. Adversarial Geospatial Abduction Problems, ACM Transactions on Intelligent Systems and Technology (02 2012)
03/09/2012	2.00	Yingqian Zhang, Efrat Manisterski, Sarit Kraus, V.S. Subrahmanian, David Peleg. Computing the fault tolerance of multi-agent deployment, Artificial Intelligence (11 2008)
05/01/2012	3.00	Paulo Shakarian, Austin Parker, Gerardo Simari, V.S. Subrahmanian. Annotated Probabilistic Temporal Logic, ACM Transactions of Computational Logic (08 2009)
05/01/2012	32.00	Massimiliano Albanese, Rama Chellappa, Vincenzo Moscato, Antonio Picariello, V.S. Subrahmanian, Pavan Turaga, Octavian Udrea. A Constrained Probabilistic Petri Net Framework for Human Activity Detection in Video, IEEE Transactions on Multimedia (10 2008)
07/09/2013	57.00	Matthias Broecheler, Andrea Pugliese, V.S. Subrahmanian. Efficient Multi-View Maintenance under Insertion in Huge Social Networks, ACM Transactions on the Web (01 2013)
07/09/2013	58.00	Andrea Pugliese, V.S. Subrahmanian, Christopher Thomas, Cristian Molinaro. PASS: A Parallel Activity Search System, IEEE Transactions on Knowledge & Data Engineering (04 2013)
07/09/2013	59.00	Massimiliano Albanese, Cristian Molinaro, Fabio Persia, Antonio Picariello, V.S. Subrahmanian. Discovering the Top-k Unxplained Sequences in Time-Stamped Observation Data, IEEE Transactions on Knowledge & Data Engineering (01 2013)

07/09/2013	60.00	Francesco Parisi, Amy Sliva, V.S. Subrahmanian. A Temporal Database Forecasting Algebra, International Journal of Approximate Reasoning (05 2013)
07/09/2013	53.00	Meirav Hadad, Sarit Kraus, Irith Ben-Arroyo Hartman, Avi Rosenfeld. Group planning with time constraints, Annals of Mathematics and Artificial Intelligence (06 2013)
07/09/2013	54.00	Cristian Molinaro, Amy Sliva, V.S. Subrahmanian. Super-solutions: Succinctly Representing Solutions in Abductive Annotated Probabilistic Temporal Logic, ACM Transactions of Computational Logic (03 2012)
07/09/2013	55.00	Rami Puzis, Manish Purohit, V.S. Subrahmanian. Betweenness Computation in the Single Graph Representation of Hypergraphs, Social Networks (04 2013)
07/09/2013	47.00	Avi Rosenfeld, Inon Zukerman, Amos Azaria, Sarit Kraus. Combining Psychological Models with Machine Learning, Synthese (10 2012)
07/09/2013	46.00	Avi Rosenfeld, Zevi Bareket, Claudia Goldman, Sarit Kraus, David LeBlanc, Omer Tsimhoni. Toward Adapting Cars to Their Drivers, AI Magazine (03 2012)
07/09/2013	48.00	Amos Azaria, Zinovi Rabinovich, Sarit Kraus, Claudia Goldman, Ya'akov Gal. Strategic Advice Provision in Repeated Human-Agent Interactions, AAAI 2012 (07 2012)
07/09/2013	61.00	Gerardo I. Simari, Maria Vanina Martinez, Amy Sliva, V.S. Subrahmanian. Focused most probable world computations in probabilistic logic programs, Annals of Mathematics and Artificial Intelligence (03 2012)

TOTAL: **21**

Number of Manuscripts:

Books

Received	Book
07/10/2013 62.00	Massimiliano Albanese, Marat Fayzullin, Jana Shakarian, V.S. Subrahmanian. Automated Coding of Decision Support Variables, New York: Springer, (01 2013)
07/10/2013 64.00	Gerardo I. Simari, Damon Earp, Maria Vanina Martinez, Amy Sliva, V.S. Subrahmanian. Forecasting Group-level Actions using Similarity Measures, New York: Springer, (01 2013)
07/10/2013 63.00	Amy Sliva, Gerardo Simari, Vanina Martinez, V.S. Subrahmanian. SOMA: Stochastic Opponent Modeling Agents for Forecasting Violent Behavior, New York: Springer, (01 2013)
07/10/2013 65.00	Maria Vanina Martinez, Amy Sliva, Gerardo I. Simari, V.S. Subrahmanian. Forecasting Changes in Terror Group Behavior, New York: Springer, (01 2013)
07/10/2013 66.00	John P. Dickerson, Gerardo I. Simari, V.S. Subrahmanian. Using Temporal Probabilistic Rules to Learn Group Behavior, New York: Springer, (01 2013)
07/10/2013 67.00	Gerardo I. Simari, John P. Dickerson, Amy Sliva, V.S. Subrahmanian. Policy Analytics Generation using Action Probabilistic Logic Programs, New York: Springer, (01 2013)

TOTAL: **6**

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Michael Ovelgoenne	0.37
Massimiliano Albanese	0.01
FTE Equivalent:	**0.38**
Total Number:	**2**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
V.S. Subrahmanian	0.02	
FTE Equivalent:	**0.02**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
Brittany Schuetzle	0.25
Rennie Silva	0.09
Amy Sliva	0.04
FTE Equivalent:	**0.38**
Total Number:	**3**

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Technology Transfer

ARO Grant W911NF-09-1-0206

Report submitted to ARO for the period April 2009 – August 2014

Final Report: PAGE: Policy Analytics Generation Engine

V.S. Subrahmanian

University of Maryland College Park

College Park, MD 20742.

vs@cs.umd.edu

301-405-6724

Abstract. SOMA (Stochastic Opponent Modeling Agents) are a paradigm within which it is possible to make statements of the form "If condition C is true in the environment in which a group G operates, then group G will take action A with probability in the range [L,U]". SOME has been used to model the behaviors of various terrorist groups and make forecasts about future attacks by those groups. The PAGE project uses SOMA models of group behavior in order to generate policies that achieve a desired goal. A policy is a way of changing the environment in which the group operates (subject to feasibility constraints). We develop such techniques on sequential machines and then develop a parallel framework for it. We also developed policies and methods by which a group of defensive resources (e.g. checkpoints) could be situated in a given geography in order to minimize attacks by an adversary on both static and moving targets (e.g. convoys). In addition, the project looked at the problem of computing between centrality in hypergraphs – also called Group Between-Ness Centrality. GBC describes the probability that a node or a group of nodes lie on a randomly chosen shortest path between two randomly selected nodes. We developed algorithms to scalably compute both BC and GBC in huge networks, outperforming (under certain conditions) all previous results. We also proposed the concept of "covertness centrality" which identifies how a "bad guy" can try to hide from "good guys" in a social network without being easily identified – by understanding how such "bad guys" can hid in social networks, we have learned how to identify them.

I. Scientific Progress and Accomplishments

During this project, we made important contributions in the following 5 broad areas.

Abductive Generation of Policies. Given a SOMA (Stochastic Opponent Modeling Agent) behavioral model consisting of rules of the form "When condition C is true in the environment of a terrorist group, then the group will carry out a terrorist attack of type A with probability in the range [L,U]", how can we take appropriate counter-terrrorism actions (subject to cost and other constraints) that will reduce the probability of attack to a level below a given threshold T. In this work, we first developed a formal probabilistic logic based model, together with sequential algorithms to address this problem and then developed parallel algorithms for the same.

Developing Patrolling Policies to Protect Static & Moving Assets against Attack. Given a geographical region *R* containing a given set of static assets (e.g. police stations, military operational sites) and a set of dynamic assets (e.g. convoys), how best can we deploy a set of checkpoints in order to minimize the possibility of a successful attack by an adversary? We developed techniques in order to optimally protect such assets based on a limited set of resources.

Geospatial Abduction. Geospatial abduction is the problem of taking a set of geolocated observations on the ground and predicting a set of partner locations that are causally linked to the observed locations. For instance, observed locations might be places where IED attacks took place, while partner locations might be places where weapons caches facilitating those attacks took place. We proposed the notion of geospatial abduction, developed complexity results for it, and proposed exact and heuristic algorithms to solve such problems. We developed a prototype system called SCARE to optimally predict locations of IED weapons caches. When applied to 21 months of data about IED attacks from Baghdad, we were able to predict locations on average to within 700 meters.

Understand Betweenness Centrality in Hypergraphs. A hypergraph consists of a set of nodes (of the hypergraph) and a set of subsets of those nodes. For instance, the nodes might represent terrorists, and the subsets might represent small cells and/or large sub-organizations. Between centrality (BC) in such hypergraphs aims at finding sets of nodes that are central in controlling the flow of communication between nodes in the hypergraph. We develop algorithms that are orders of magnitude faster than past work for computing BC on large hypergraphs. The results also scale up BC computation on regular graphs.

Covertness Centrality in Networks. In real world terrorist networks, the terrorist network is embedded within a much larger network composed largely of ordinary people. How do terrorists stay hidden in such networks, given that they need to communicate with one another while staying "below the radar"? We present the fundamentally new notion of covertness centrality that achieves this purpose.

II. CONTRIBUTIONS

II.A Abductive Generation of Policies

In past work, we showed that a class of probabilistic logic programs called action probabilistic logic programs (ap-programs for short) could be derived automatically from a body of terrorism data in order to build models of the behaviors of over 40 terror groups from Morocco to Afghanistan. In ap-logic programs, all predicate symbols are categorized either as action predicate symbols or environmental predicate symbols. ap-logic programs consist of a finite set of rules of the form "if condition C is true in the environment in which group G operates, then group G will take action A with a probability in the range [L,U]". ap-logic programs are augmented with a ``state'' (set of facts about the current environment in which the group operates). Only environmental predicate symbols can appear in the body of a rule, and only action predicate symbols can occur in the head of a rule. The state is a set of ground atoms whose associated predicate symbols are all environmental predicate symbols.

Suppose, rather than predicting what action(s) a group would take in a given situation or environment, we want to determine what we can do in order to induce a given behavior by the group. Furthermore, suppose there are some constraints limiting what we can do. Certain things might be do-able, while others might be impossible. Given a specification of various actions that we can actually take, and given a desired behavior G that we wish to elicit from the group, and given that we want the probability of behavior G to be in the probability interval [L,U], how can we change the state so that the group is predicted to exhibit behavior G with a probability in the interval [L,U]? We call this the probabilistic logic abduction problem (PLAP).

We first showed that the basic PLAP problem is EXPTIME-complete. This means that any "exact" solution to the Basic PLAP problem is likely to take an exponential amount of time. For the sake of comparisons, we developed an algorithm that computed this exact solution, even though we knew it would take a lot of time to run. We then developed a more efficient algorithm that used the notion of a state-reachability graph to reduce the search space. We then came up with a very efficient algorithm to solve the Basic PLAP problem in the case when either L=0 or U=1. In fact, these two cases are the ones most likely to be of practical interest because users are usually interested in finding a way to change the state so that the probability that the group engages in a desired behavior is greater than or equal to some threshold (corresponding to the U=1 case) or they are interested in finding a way to change the state so the probability that the group engages in an undesired behavior is below some threshold (corresponding to the case when L=0). We showed that as long as we make these assumptions, a much more efficient algorithm is possible and we went ahead and developed one. We showed this algorithm is correct under the stated assumptions. However, this only seems to reduce the problem from being EXPTIME-complete to NP-hard. We extended this to a host of complexity results for PLAP under varying assumptions.

We developed a prototype implementation and experiments showing that our algorithm is feasible to use even when the *ap*-program contains hundreds of rules.

Following this, we took the problem one step further by reasoning about how the entity being modeled reacts to our efforts. We are interested in identifying the best course of action on our part, given some additional inputs regarding the cost of exerting influence in the environment and how desirable certain outcomes are; this is called the *cost-based abduction problem* (CBA). We then investigate an approach to solving this problem exactly based on Markov Decision Processes, showing that this approach quickly becomes infeasible in practice. Afterwards, we describe a novel heuristic algorithm based on probability density estimation techniques that can be used to tackle CBA with much larger instances. We then develop the first parallel algorithms for abduction in probabilistic logics.

Our prototype implementation and experimental results show that our parallel algorithm scales well in practice and achieves results that are useful in practice.

II.B Developing Patrolling Policies

In this work, we consider two problems related to the protection of assets in a road network. The first problem assumes that certain arbitrary vertices (denoting assets) in a graph (representing the road network) must be protected from adversaries who may be located at any subset of vertices. We call this the static asset protection problem (SAP) because the asset being protected is static. For example, the police in a US city may be protecting a hotel where a famous politician is staying for a few days. In contrast, the dynamic asset protection problem (DAP, for short) considers the case where the asset being protected is moving along pre-determined route. For instance, a politician may be traveling along a parade route and the police need to protect the entire route. In both cases, the police have limited resources to protect the assets in question.

Both problems areintimately related to network interdiction [9, 17], where an enemy attempts to traverse a graph from a start vertex to an end vertex while an interdictor impedes his progress by "breaking" edges in the graph. Work in network interdiction has traditionally focused on stopping enemy movement along some path; however, our work is motivated by a need to protect a static asset's position or dynamic asset's path.

In this work, we first formalize the static asset protection problem and define an "optimal" deployment of resources to protect the asset in question, taking adversarial behavior into account. We subsequently develop a formal theoretical model based on minimal edge cuts in graphs and show that randomization over what we call single minimal edge cuts yields the optimal asset protection. We subsequently propose an algorithm for SAP and analyze its running time showing that it works well.

We then define the dynamic asset protection problem and show that this problem is NP-complete. We propose a greedy algorithm that tries to quickly compute a (sub-optimal) way of solving DAP.

We developed a prototype implementation of tall of these algorithms and conducted extensive experiments using road networks drawn from real cities. These results show that our algorithm performs very well on real world data for both SAP and DAP.

II.C Geospatial Abduction Problems

In addition to our work on abduction in the context of ap-programs, we consider the following problem. Given a set *OBS* of events on the ground and given some related locations *LOC* that we want to infer from the observations, how best can we infer the latter from the former. For instance, OBS could consist of locations where IED attacks took place, while LOC could consist of locations of weapons caches supporting those attacks.

In order to solve such problems, we formally defined a *geospatial abduction problem* as one consisting of the following inputs

- A map M

- A set OBS of (x,y)-coordinates denoting places where observations were made

- Two numbers L and U such that 0 < L < U.

- A feasibility predicate that takes each point on the map M and assigns it either "true" or "false"

- An integer k > 0

Intuitively, the feasibility predicate returns "yes" for a point (x,y) if and only if (x,y) satisfies some conditions for the location we want to identify. For instance, in the case of Shiite-backed IED attacks, we might set the feasibility predicate to return "true" on a location (x,y) iff:

• (x,y) is not in a Sunni neighborhood and

• (x,y) is not in a body of water (e.g. the Tigris river)

The two numbers L and U can be learned from historical data. L denotes a lower bound on the distance between an IED cache location and an attack location, while U denotes an upper bound on the same distance.

The Geospatial Abduction Problem is to find a set of at most *k* locations which best "explain" the observations being seen. For each observation, one of the *k* locations returned by the algorithm must be no less than L units of distance from the observation and no more than U units of distance. Moreover, each of the *k* locations returned must be feasible according to the feasibility predicate. In our work, we learned L and U from a historical collection of data. We showed that the Geospatial Abduction Problem is computationally intractable and is NP-complete. We also proved that several variants of the problem are also NP-hard. This makes it unlikely that this problem can be solved exactly.

As a consequence, we developed algorithms to solve the problem approximately. To do so, we showed that geospatial abduction problems can be reduced to classical combinatorial optimization problems such as set cover and dominating set in polynomial time. As a consequence, we can leverage approximation methods for these problems in designing a solution to geospatial abduction problems. We defined several such approximation algorithms. We developed a prototype system called SCARE and showed that the best algorithm for our problem is very efficient and we showed – using 21 months of data about IED attacks in Baghdad – that it can accurately pinpoint weapons caches in Baghdad to within 0.45 miles of its location. (The test data set was 14 months of data, while 7 months of data was used for training). A screenshot of SCARE is shown below.

SCARE has since been transitioned to many DoD units including units in Baghdad. Entities to which SCARE was transitioned include ARL, FORSCOM, NCTC, FBI/Interpol.

II.D Understanding Between-ness Centrality in Hypergraphs

Consider a social network N = (V,E,w) consisting of a set V of vertices and a set E of edges. w is a function that assigns weights to the edges in E. Thus, if w(u,v) is high, then this means that the strength of the relationship between u and v is high.

For instance, V might be the set of all nodes in a known terror network, and w(u,v) might denote the number of phone calls we have intercepted from u to v. In such an application, we are interested in finding sets of nodes which – if "subverted" – would significantly disrupt communications within the terror network. We would obviously like to a find a set G of nodes (or a "group" of nodes) whose subversion, in this example, would maximize the "disruption".

If we ignore our putative example scenario, in general, we would like to identify the “joint centrality” of a set G of vertices in the social network N = (V,E,w).

Group between-ness centrality of G is, intuitively, a heuristic estimation for the *percentage* of messages flowing in the network that pass through at least one vertex in G. For instance, if a group G has high group between-ness centrality, then this means that ”subverting” all vertices in G will lead to significant communications reduction for the adversary.

We can formally define group between-ness centrality (GBC) as follows. Given two arbitrary vertices *s,t* in V, we first define a quantity called GBC(G,s,t) as follows. Let E(s,t) be the expected number of messages flowing from s to t.

$$GBC(G,s,t) = \frac{\#\ of\ shortest\ paths\ from\ s\ to\ t\ in\ N\ traversing\ at\ least\ one\ vertex\ in\ G}{number\ of\ shortest\ paths\ from\ s\ to\ t\ in\ N}$$

The group between-ness centrality of group G can now be defined as:

$$GBC(G) = \sum_{s \neq t} GBC(G,s,t)$$

In other words, the group between-ness centrality of G aggregates the group-betweenness centrality across all pairs of distinct vertices in N.

We first focused on scaling the computation of between-ness centrality (BC) for undirected graphs. Between-ness centrality is the special case of GBC when the size of G is 1. In order to scale BC computation, we built on top of the classical Brandes algorithm which has two phases – a *forward phase* in which we calculate the number of shortest paths going through a given node, and a *backward phase* where we aggregate these numbers to compute a between-ness centrality value.

Many graphs naturally contain “cliques”. Simply put, a clique C of a graph G=(V,E) is a set of nodes such that for any pair of vertices, v1,v2 in C, there is an edge between v1 and v2 linking the two. The principal idea within the algorithm we developed is the following. We asked ourselves the question: Can we somehow merge all nodes in a clique into a single “supernode”. By doing so, we replace graph G by a super-graph SG whose vertices are the cliques of G. When there are lots of large cliques in G, this naturally reduces the size of G significantly.

Once we consider the super-graph SG, we can compute between-ness of the “super-nodes” in SG using a classical algorithm like Brandes’ algorithm or in fact any other classical BC algorithm.

We then generalized these results to apply to hypergraphs. There are numerous real-world networks that can be viewed naturally as hypergraphs. Here are a few simple examples.

• In *FaceBook*, *LinkedIn* and *YouTube*, all members of any of the groups one can join, can be viewed as a single hyperedge.

• In *LinkedIn*, all people who studied at the same university (or work at the same company) can be viewed as a single hyperedge.

• In an online citation network like *CiteSeer* or *DBLP*, all people who publish in the same conference may be viewed as a single hyperedge.

• In *Flickr* all people who like beach pictures or gorilla pictures might form a hyperedge.

• In a movie database like *IMDB*, we might consider everyone who acted in the same movie to constitute a hyperedge.

Many of these examples correspond to bipartite graphs (e.g. person vertices and group vertices in *FaceBook*, *LinkedIn* and *YouTube* above; person vertices and university vertices in *LinkedIn* above, people vertices and conference vertices in *CiteSeer* and *DBLP* above). Such graphs are also called two-mode graphs and they are known to have a one-mode projection corresponding exactly to a hypergraph as defined above. For instance, all members of the same *LinkedIn* group form a hyperedge as do all alumni of the same university in *LinkedIn*.

Thus, we see that many online social networks, which are traditionally represented as undirected graphs, can also be viewed as hypergraphs, with membership in a common group or community as the property that determines the set of hyperedges.

We study the following question: how can we significantly speed up the exact computation of BC on very large graphs? We show that the one mode representation of a bipartite graph can enable us to significantly speed up computation in such graphs, even when the graphs are huge and have millions of vertices. The basic idea of our work is that betweenness centrality computation for the simple graph representation of hypergraphs can be faster due to the nature of that representation. The simple graph representation results in a graph with many cliques, and many of the nodes in such cliques can be "aggregated" in BC computation as their influence on the number of shortest paths between pairs can be easily computed. We build upon these past results in order to develop methods to exactly compute betweenness centrality of vertices in very large hypergraphs (using the one mode representation). Specifically, we make the following contributions:

- We first develop a naive adaptation (*NaiveHBC*) of the traditional BC algorithm due to Brandes and show that it can also be used to compute *HBC*.
- We then develop a *HyperBFS* algorithm that improves the "forward phase" of the BC algorithmfor hypergraphs. Thus, our algorithm may be viewed as an extension of Brandes' classical BC algorithm to the case of hypergraphs with some novel new improvements both to the forward search and backpropagation parts, as well as new notions to reduce graph size.
- In order to reduce graph size without any loss of accuracy, we develop the concept of "bridging" and "non-bridging" vertices in a hypergraph and use these concepts to associate a *compound hypergraph* with any hypergraph. We show that the betweenness centrality of nodes in the original hypergaph can be easily computed from the compound hypergraph.

- Next, we developed a new *HyperBC* algorithm for computing *HBC* in hypergraphs by using *HyperBFS* and by a more sophisticated backward pairwise accumulation step. Note that *HyperBC* is an *exact and sequential* BC computational algorithm for hypergraphs, not an approximation algorithm or a parallel algorithm.
- We report on the results of detailed experiments on 5 real-world data sets — two of these are based on *DBLP* (but with different definitions of what the hypergraphs are), and one each on *IMDB*, *FlickR* and *TheMarker*. In addition, we report results on synthetic data sets. Our experiments show that our algorithms can improve performance of *NaiveHBC* by 1 to 3 orders of magnitude. The speedups improve as the size of hyperedges increase. We further show that most of the speedups are due to the *HyperBFS* algorithm and the concept of bridging/non-bridging vertices, and not because of the backpropagation phase of the algorithm. We test our algorithms on several real and synthetic data sets including the *IMDB − Full* dataset that contains 2.89M vertices and show that our algorithms provide reasonable performance in all of these cases — in comparison, the *NaiveHBC* algorithm does not work in such large cases. We also compared our algorithm with recent algorithms and show that except in the case of small networks, our algorithm outperforms those algorithms are well.

II.E Covertness Centrality

Social networking research has focused extensively on the problem of identifying important vertices in a network, taking into account the structural properties of the network. As a consequence, important concepts such as degree centrality, between-ness centrality, closeness centrality, and eigenvector centrality have been proposed (as well as many others that we do not list here). Work in this arena dates back to the beginning of the last century.

More recently, it has been asserted that such structural centrality measures can be used to identify important players in a network, e.g. leaders of a drug ring or key players in a terrorist network. In this work, we asked ourselves the question: what happens if a drug kingpin or a terrorist leader needs to communicate with a set *I* of individuals in a network (e.g. *I* could consist of people known to law enforcement or intelligent authorities who are not aware that they have been uncovered) and if he is smart enough to know that measures such as classical centrality measures are used against him? In this case, we he would take measures to ensure that he does not have high centrality in any of the set *C* of measures that he thinks might be used to identify him. Nevertheless, he must be able to communicate with the individuals in *I.*

In order to identify "intelligent" adversaries who wish to remain covert in the presence of an agent seeking to uncover them using a set *C* of classical centrality measures, we developed the concept of *covertness centrality* that consists of two key parts:

- *Common-ness:* Informally speaking, the common-ness of a vertex is the percentage of vertices in the network that have a similar centrality according to the measures in *C*. Though we can measure common-ness in many different ways, in this work we chose not to do so – rather we *first* identified a set of axioms that we believed *any* reasonable common-ness measure should satisfy and then we identified two specific common-ness measures, each with some advantages.

- *Communication Ability.* The communication ability of a vertex was defined in terms of the connectivity of that vertex to each of the vertices in *I* that the vertex being considered needs to communicate with.

Based on these two concepts, we defined covertness centrality. We developed two classes of sampling based algorithms to compute covertness centrality of vertices on 3 real-world datasets – one consisting of emails (small), a 40K node YouTube data set, and a 60K node YouTube data set. Our algorithms scaled well, giving good performance as shown in the table below.

	URV	Youtube 40k	Youtube 60k
Computing time	0.1second	2 seconds	3 seconds

Moreover, even though we used sampling based approaches, we computed the Kendall Rank Correlation Coefficient to test whether the top *k* covert nodes identified by our algorithm agreed with the top *k* identified without sampling. As the table below shows, the results were very strongly correlated with a Kendall's tau-coefficient over 0.9

Network	Simple			Systematic		
	20%	40%	60%	20%	40%	60%
URV	0.965	0.981	0.981	0.966	0.985	1.000
YouTube 40k	0.975	0.992	0.997	0.979	0.996	1.000
YouTube 60k	0.981	0.995	0.997	0.990	0.996	1.000

III. STUDENTS/POSTDOCS INVOLVED

Students: J. Dickerson, C.Kang, M. Nagel, A. Parker, M. Purohit, A. Sawant, B. Shuetzle, P. Shakarian, G. Simari, A. Sliva

Postdocs: C. Molinaro, M. Ovelgonne, R. Puzis

PAPERS PUBLISHED/ACCEPTED/SUBMITTED

1. M. Ovelgonne, C. Kang, A. Sawant and V.S. Subrahmanian. Covertness Centrality in Networks, accepted for publication in Proc. 2012 Intl. Symposium on Foundations of Open Source Intelligence and Security Informatics (FOSINT-SI), Istanbul, Turkey, August 2012.
2. P. Shakarian#, M. Nagel#, B. Schuetzle#, and V.S. Subrahmanian^. Abductive Inference for Combat: Using SCARE-S2 to Find High-Value Targets in Afghanistan, *Proc. 2011 Intl. Conf. on Innovative Applications of Artificial Intelligence*, Aug. 2011, AAAI Press.
3. G. Simari#, J. Dickerson#, A. Sliva# and V.S. Subrahmanian^. "Parallel Abductive Query Answering in Probabilistic Logic Programs", ACM Transactions on Computational Logic, 14(2):12, 2013.
4. G. Simari# and V.S. Subrahmanian^. Abductive Inference in Probabilistic Logic Programs, Tech. Communications of the 2010 Intl. Conf. on Logic Programming, Edinburgh, UK, July 2010, pp. 192-201.
5. P. Shakarian#, V.S. Subrahmanian^ and M.L. Sapino. GAPS: Geospatial Abduction Problems, ACM Transactions on Intelligent Systems and Technology, Vol. 3, Nr. 1, Oct 2011, Article Nr. 7, doi>10.1145/2036264.2036271
6. G. Simari#, A. Sliva#, A. Parker#, V.S. Subrahmanian^. Approximate Achievability in Event Databases, Proc. 11th European Conf. on Symbolic & Quantitative Approaches to Reasoning with Uncertainty, June 2011, Ireland, pp. 737-748.
7. P. Shakarian#, A. Parker#, G. Simari# and V.S. Subrahmanian^. Annotated Probabilistic Temporal Logic, ACM Transactions on Computational Logic, Vol. 12, Nr. 2, 2011.
8. M. Broecheler, A. Pugliese, and V.S. Subrahmanian. Efficient Multi-View Maintenance in the Social Semantic Web, WWW (Companion Volume) 2012.
9. C. Kang#, J. Grant, A. Pugliese, and V.S. Subrahmanian^. STUN: Spatio-Temporal Uncertain (Social) Networks, Proc. 2012 International Conference on Advances in Social Network Analysis and Mining (ASONAM 2012), August 2012, Istanbul, Turkey (full paper – 16% acceptance rate).
10. C. Kang#, A. Pugliese, J. Grant, and V.S. Subrahmanian^. STUN: Querying Spatio-temporal Uncertain (social) Networks. Social Network Analysis and Mining journal 4(1) (2014).
11. C. Kang, C. Molinaro, S. Kraus, Y. Shavitt, and V.S. Subrahmanian. Diffusion Centrality in Social Networks**,** accepted for publication at 2012 International Conference on Advances in Social Network Analysis and Mining (ASONAM 2012), August 2012, Istanbul, Turkey (short paper – 25% acceptance rate).
12. Puzis, Rami, Manish Purohit, and V. S. Subrahmanian. "Betweenness computation in the single graph representation of hypergraphs." *Social Networks* 35.4 (2013): 561-572.
13. P. Shakarian# and V.S. Subrahmanian^. Region-based Geospatial Abduction with Counter-IED Applications, in: Counterterrorism and Open Source Intelligence (ed. U.K. Wiil), Springer Verlag Lecture Notes on Social Networks, pp. 105-142.