Spatio-Temporal Nonlinear Filtering with Applications to Information Assurance and Counter Terrorism

Boris Rozovsky

Brown University

8 PUBLICATIONS **26** CITATIONS

SEE PROFILE

Alexander G. Tartakovsky

University of Connecticut

100 PUBLICATIONS **1,649** CITATIONS

SEE PROFILE

Aram Galstyan

University of Southern California

99 PUBLICATIONS **1,660** CITATIONS

SEE PROFILE

Gérard G. Medioni

University of Southern California

383 PUBLICATIONS **11,497** CITATIONS

SEE PROFILE

REPORT DOCUMENTATION PAGE		Form Approved OMB NO. 0704-0188	

Public Reporting Burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington DC 20503

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE:	3. REPORT TYPE AND DATES COVERED Final Report 1-May-2006 - 30-Sep-2006	
4. TITLE AND SUBTITLE Spatio-Temporal Nonlinear Filtering with Applications to Information Assurance and Counter Terrorism		5. FUNDING NUMBERS W911NF-06-1-0094	
6. AUTHORS A. Galstyan, A. Bertozzi, P. Cohen, G. Medioni, C. Papadopolous, B. Rozovsky, A. Tartakovsky, and V. Veeravalli		8. PERFORMING ORGANIZATION REPORT NUMBER	
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES University of Southern California Dept. of Contracts & Grants 837 W. 36th Place, STO 330 Los Angeles, CA 90089 -1147			
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. SPONSORING / MONITORING AGENCY REPORT NUMBER 50363-MA-MUR.1	
11. 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.			
12. DISTRIBUTION AVAILIBILITY STATEMENT Approved for Public Release; Distribution Unlimited		12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The abstract is below since many authors do not follow the 200 word limit			
14. SUBJECT TERMS Optimal Spatio-Temporal Nonlinear Filtering; Change Detection; Information Fusion; Hats Simulator; Tracking Terrorists; Video Tracking Extended Objects			15. NUMBER OF PAGES Unknown due to possible attachments 16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION ON THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev .2-89)
Prescribed by ANSI Std.
239-18 298-102

Report Title

Spatio-Temporal Nonlinear Filtering With Applications to Information Assurance and Counter Terrorism

ABSTRACT

The objective of this MURI project is to develop a general and systematic foundation and algorithms for spatial-temporal statistical inference and for fusion of heterogeneous information from multi-source, multi-sensor distributed sensor networks. Immediate applications of the proposed work are Network Centric Warfare, where new and emerging systems such as MASINT and FORCENet collect but do not adequately interpret vast amounts of data; and homeland security applications, including video monitoring, and near-field and far-field intelligence analysis. Our research will solve three central problems: (a) nonstationarity, (b) integrating metric and symbolic information, and (c) very high dimensionality. Current methods for pattern recognition in monitoring and surveillance are designed for stationary patterns, and cannot cope with new patterns in ever-changing environments. We develop new statistical methods for the nonstationary environment, particularly spatio-temporal nonlinear filtering, change-point detection, and advanced fusion methods. A distinctive feature of our approach is that the spaces in which estimation, classification and tracking is performed are both metric and symbolic. Just as a moving vehicle may be tracked in a metric coordinate space by conventional filters, so can an unfolding terrorist plan be tracked in plan space by a hybrid metric-symbolic nonlinear filter.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

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

See the attached file

Number of Papers published in peer-reviewed journals: 1.00

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

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

Number of Presentations: 0.00

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

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

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

See the attached file

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 8

(d) Manuscripts

See the attched file

Number of Manuscripts: 5.00

Number of Inventions:

Graduate Students

NAME	PERCENT_SUPPORTED	
Included in the attached pdf file		No
FTE Equivalent:		
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED	
Included in the attached pdf file		No
FTE Equivalent:		
Total Number:	**1**	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Included in the attached pdf file		No
FTE Equivalent:		
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	
Included in the attached pdf file		No
FTE Equivalent:		
Total Number:	**1**	

Names of Personnel receiving masters degrees

NAME	
Included in the attached pdf file	No
Total Number:	**1**

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

1 a. University of Illinois, Urbana-Champaign

1 b. Electrical and Computer Engineering De
128 CSRL, 1308 West Main Street
Urbana IL 61801

Sub Contractor Numbers (c):

Patent Clause Number (d-1):

Patent Date (d-2):

Work Description (e):

Sub Contract Award Date (f-1):

Sub Contract Est Completion Date(f-2):

---

1 a. University of California, Los Angeles

1 b. Department of Mathematics
520 Portola Plaza
Los Angeles, CA CA 90095

Sub Contractor Numbers (c):

Patent Clause Number (d-1):

Patent Date (d-2):

Work Description (e):

Sub Contract Award Date (f-1):

Sub Contract Est Completion Date(f-2):

Inventions (DD882)

SPATIO-TEMPORAL NONLINEAR FILTERING WITH APPLICATIONS TO INFORMATION ASSURANCE AND COUNTER TERRORISM

FINAL PROGRESS AND TECHNICAL REPORT

GRANT # W911NF-06-1-0094

REPORT # FPTR-CAMS-1-06

DATES COVERED: 05/01/2006–09/30/2006

AUTHORS:
A. Bertozzi, P. Cohen, A. Galstyan, G. Medioni,
C. Papadopolous, B. Rozovsky, A. Tartakovsky, and V. Veeravalli

LEADING PERFORMING ORGANIZATION:

Center for Applied Mathematical Sciences
University of Southern California
1042 Downey Way, DRB-308
Los Angeles, CA 90089

Principal Investigator: Boris Rozovsky

Co-Principal Investigator: Alexander Tartakovsky

SPONSORING/MONITORING AGENCY:

The U.S. Army Research Office
Mathematical Sciences Division
P.O. Box 12211
RTP, NC 27709

POC: Dr. Mou-Hsiung Chang

Contents

1. SCIENTIFIC PERSONNEL SUPPORTED BY THIS PROJECT

University of Southern California (USC):

1. Gen Bartlett, PhD student, USC-CS
2. Paul Cohen, Director, USC-ISI
3. Aram Galstyan, Research Associate, USC-ISI
4. Sinjini Mitra, Postdoc, USC-ISI
5. Clayton Morrison, Postdoc, USC-ISI
6. Christos Papadopoulos, Assistant Professor, USC-CS
7. Andrew Papanicolaou, PhD student, USC-CAMS-DM
8. Alexey Polunchenko, PhD student, USC-CAMS-DM
9. Boris Rozovsky, Professor, USC-CAMS-DM
10. Alexander Tartakovsky, Associate Director, Research Professor, USC-CAMS-DM
11. Gerard Medioni, Professor, Dept Chair, USC-CS

University of Illinois at Urbana-Champaign (UIUC):

1. Venugopal Veeravalli, Professor, UIUC-ECE
2. Sahand Ahmad, PhD student, UIUC-ECE

University of California at Los Angeles (UCLA):

1. Andrea Bertozzi, Professor of Mathematics
2. Tony Chan, Professor of Mathematics
3. P. Jeffrey Brantingham, Assistant Professor of Anthropology
4. Maria D’Orsogna, Assistant Researcher, Department of Mathematics
5. Yuan R. Huang, Masters student in Electrical Engineering
6. Yao-Li Chuang, Physics PhD student, Duke University
7. Mi Youn Jung, PhD student in Mathematics, UCLA
8. Zhipu Jin, Postdoc in Mathematics, UCLA
9. Vlad Voroninsky, Undergraduate in Applied Mathematics and Cybernetics, UCLA
10. Abhijeet Joshi, Undergraduate in Electrical Engineering
11. Kevin Leung, Masters student in Electrical Engineering, UC Davis

2. SUMMARY OF ADDRESSED TASKS AND ACCOMPLISHMENTS

2.1. List of Publications

1. J. von Breach, S.R. Thiruvenkadam, and T. Chan, "Occlusion Tracking with Logical Models," preprint.
2. Y.-L. Chuang, Y.R. Huang, M.R. D'Orsogna, and A.L. Bertozzi, "Multi-Vehicle Flocking: Scalability of Cooperative Control Algorithms Using Pairwise Potentials," accepted in *The 2007 IEEE International Conference on Robotics and Automation*, 2007.
3. J. Cvitanic, R. Liptser, and B. Rozovskii, "A Filtering Approach to Tracking Volatility from Prices Observed at Random Times," *Annals of Applied Probability*, vol. 16, no. 3, pp. 1633-1652, 2006.
4. J. Cvitanic, B. Rozovskii, and I. Zalyapin, "Numerical Estimation of Volatility Values from Discretely Observed Diffusion Data," *J. Numerical Fin.*, accepted in 2006.
5. J. Pummeled and V.V. Veeravalli, "Smart Sleeping Policies for Energy Efficient Tracking in Sensor Networks," Submitted to the *IEEE Transactions on Signal Processing*, October 2006.
6. J. Pummeled and V.V. Veeravalli, "Smart Sleeping Strategies for Localization and Tracking in Sensor Networks," *Proceedings of the 40th Allobar Conference on Signals, Systems, and Computers*, Monterey, CA, November 2006. (**Invited**)
7. A. Hussain, J. Heidemann, and C. Papadopoulos, "Identification of Repeated Denial of Service Attacks," *Proceedings of IEEE Infocom 2006*, Barcelona, Spain, 2006.
8. Y. Landa, D. Galkowski, Y.R. Huang, A. Joshi, C. Lee, K.K. Leung, G. Malla, J. Treanor, V. Voroninski, A.L. Bertozzi, and R. Tsai, "Robotic Path Planning and Visibility With Limited Sensor Data," to appear in *The 2007 American Control Conference*, 2007.
9. K.K. Leung, C.H. Hsieh, Y.R. Huang, A. Joshi, V. Voroninski, and A.L. Bertozzi, "A Second Generation Micro-Vehicle Testbed for Cooperative Control and Sensing Strategies," to appear in *The 2007 American Control Conference*, 2007.
10. M. Pollak and A.G. Tartakovsky, "On Asymptotic Exponentiality of the Distribution of First Exist Times for a Class of Markov Processes," Submitted to *The Annals of Applied Probability*, September 2006.
11. A.G. Tartakovsky and H. Kim, "Performance of Certain Decentralized Distributed Change Detection Procedures," *Proc. 9th International Conference on Information Fusion*, Florence, Italy, 10-13 July 2006, CD ISBN 0-9721844-6-5, IEEE Catalog No. 06EX1311C.
12. A.G. Tartakovsky, "Asymptotic Optimality in Bayesian Change-Point Detection Problems Under Global False Alarm Probability Constraint," Submitted to *Theory of Probability and its Applications*, August 2006.
13. A.G. Tartakovsky, "An Asymptotically Optimal Change Detection Strategy Under Nontraditional Global False Alarm Probability Constraint," *The 2007 Taipei International Statistical Symposium and ICSA International Conference*, Taipei, Taiwan, 24-28 June 2007. (**Invited**)
14. Q. Yu, I. Cohen, G. Medioni and B. Wu, "Boosted Markov Chain Monte Carlo Data Association for Multiple Target Detection and Tracking," *Proceedings of the 18th International Conference On Pattern Recognition (ICPR06)*, vol. 2, pp. 675-678, 2006.

2.2. Summary of Objectives and Accomplishments

During this Phase 1 of our research we have directed our efforts towards identifying and formulating the key research problems pertaining to this project.

Main objectives and accomplishments:

- *Efficient Bayesian Approach for Recognition of Patterns and Trends – Spatio-temporal Nonlinear Filtering.* Our main objective is to develop new nonlinear filtering algorithms that rely on both spatial and temporal information. Development of this technology will provide tools that can analyze simultaneously large number of targets or complicated evolving spatial patterns based on incomplete noisy observations. The analysis and synthesis of high-volume data and, in particular, spatio-temporal data is addressed by different disciplines and from different perspectives. Our approach is probabilistic in nature, more specifically it is Bayesian. One important feature of the Bayesian approach is that it interprets the data not as a self-contained information depository but in the light of already available knowledge (e.g. human expertise) regarding the events reflected by the data. This feature is an ideal instrument for keeping human operators "in the loop" in the process of automated decision making. Nonlinear filtering (NLF) is an extension of the Bayesian framework into dynamical systems. NLF is a field on the cutting edge of contemporary stochastic analysis, information theory, and statistical inference. This is an emerging methodology with enormous breadth of applications. An important subset of NLF algorithms developed for Markovian dynamics is often referred to as Hidden Markov Models (HMM). The outputs of a NLF are sequentially computed estimates (posterior distributions) of the states of evolving hidden/noisy patterns. Accomplishments to date:
  - In the first phase of our research, we have directed our efforts towards identifying the approaches to nonlinear filtering that are most suitable in dealing with spatial-temporal data sets.
  - We have extended the Wiener Chaos methodology to non-causal dynamical and stationary infinite-dimensional systems that are modeling complicated behavior of multiple agents.
- *Multiple Target Detection and Tracking from a Moving Platform.* Detecting and tracking multiple target is a critical component of video surveillance, as it provides the description of spatio-temporal relationships among moving objects in the scene required by activity recognition modules. Environments of interest usually contain an unknown and varying number of moving targets which enter and exit the field of view randomly, and might remain within the field of view during the whole sequence. Automatic detection and tracking of multiple targets involves the detection of moving regions, the initialization of tracks, the association of regions across time and the filtering of erroneous detections or tracks. Instead of separating the detection and tracking as two separate procedures, we propose a probabilistic framework for automatic detection and tracking of objects, which combines the detection and tracking together. This allows object detection to make use of temporal consistency and facilitates robust tracking of the object. Moreover, we formulate the multiple targets tracking as a data

association problem, in which the purpose is to find the best association between observations (i.e., detected moving regions) and targets while maximizing the posterior association probability.

- *A Scalable Framework for Identifying and Tracking Covert Activities of Hostile Agents.* This framework will include a family of algorithms for tracking/monitoring covert plans based on noisy observations, and a theoretical methodology for analyzing these algorithms. We use an approach based on probabilistic models for tracking collaborative plans based on Hierarchical Hidden Markov Model and its extensions. These algorithms will be validated on data from the Hats simulator, which is a lightweight proxy for many intelligence analysis problems.

- *Pattern Change and Trend Detection in Distributed Multisensor Systems With Applications to Network Security and Surveillance.* The overarching goal of this part of the project is to develop new procedures for change detection in distributed multisensor systems, and to provide an analytical framework to predict their performance in terms of the tradeoff between detection delay and frequency of false alarms. To address this goal, we propose to analyze several generalizations of the change detection problem that arise in the applications to distributed sensor systems. We consider the configuration where the sensors communicate to a common fusion center. The change in the statistics of the observations at the sensors is governed by the event. We investigate a variety of models for the change process: only one (or a subset) of the sensors changes, they all change at the same time, or they change at different times. We also include various scenarios for communication with the fusion center, from the centralized one where the sensors send sufficient statistics, to the decentralized one where they send quantized observations or local decisions. We study the role of feedback from the fusion center, and investigate schemes for conserving energy at the sensors such as switching the sensors between on/off modes and censoring their observations. Our strategy for design and analysis accommodate general statistical models for the observations, and allow for different degrees of model uncertainty. Specific objectives and accomplishments are:

  - Development of an efficient Bayesian approach for recognition of patterns and trends, including joint nonlinear filtering (NLF) and hypothesis testing methods.
  - Pattern change and trend detection in distributed sensor systems, in particular optimal adaptive parametric and nonparametric change-point detection procedures and information integration and decision fusion in distributed heterogeneous multi-source multi-sensor systems.
  - Intrusion detection and networking problems in information assurance (IA), in particular implementation of advanced statistical methods for designing a scalable forensic Intrusion Detection System (IDS) and development of an adaptive hierarchical IDS.
  - Results include: (a) Preliminary formulation of the problems in decentralized distributed sensor systems, (b) Initial results on asymptotically optimal and suboptimal change-point detection in distributed sensor systems, (c) Initial results on distributed IDS design and detection of computer intrusions, (d) Results on target detection and tracking in heavy clutter.

- *Energy-Efficient Tracking in Sensor Networks.* We study the problem of tracking an object that is moving randomly through a dense network of wireless sensors. We assume that each sensor has a limited range for detecting the presence of the object, and that the network is sufficiently dense so that the sensors cover the area of interest. In order to conserve energy the sensors may be put into a sleep mode with a timer that determines the sleep duration. We assume that a sensor that is asleep cannot be communicated with or woken up. Thus the sleep duration needs to be determined at the time the sensor goes to sleep based on all the information available to the sensor. The objective is to track the location of the object to within the accuracy of the range of the sensor. However, having sleeping sensors in the network could result in tracking errors, and hence there is a tradeoff between the energy savings and the tracking errors that result from the sleeping actions at the sensors. We consider the design of sleeping policies that optimize this tradeoff.

- *Spatio-Temporal Image Segmentation and Video Tracking Using Logical Models.* We designed and built a second generation robotics testbed with onboard computing and onboard sensing. The developed algorithm for tracking under partial occlusions utilizes Logic Models with the addition of prior shape information. We represent object motion as a registration between frames. We can track successfully as long as the object of interest maintains nearly constant shape and intensity throughout the sequence, and does not become totally occluded.

- *Cooperative Control Algorithms for UCLA Multivehicle Wireless Testbed.* We develop information fusion algorithms for agile or mobile sensors with improved performance by linking current deterministic methods with stochastic NLF/HMM and change-point detection approaches.

- *Models for Spatio-Temporal Dynamics of Criminal Behavior.* Real-time integration of information from the variety of surveillance and sensor platforms. Multiple system platforms include video-surveillance, distributed environmental sensing, and event recognition and patterns from law enforcement agencies.

- *Spectral Analysis Techniques for Real-time Generation of Signatures of Computer Network Attacks*. This objective targets development of novel Spectral Analysis Techniques to generate signatures of attacks that cannot be detected with current IDSs. These include encrypted attacks, low level attacks and attacks through proxies. An encrypted attack is an attack where the packet stream is encrypted and the IDS cannot read the application headers or payload. A low-level attack is one where rather than use relatively few zombies attacking at full speed, many more zombies are used, each attacking with just a few packets per second in order to stay below the IDS threshold of an attack. Finally, an attack through a proxy is one where malicious and legitimate packets through a proxy become indistinguishable to an IDS and the only way to stop the attack is to filter all packets through the proxy. We have preliminary work showing that we can create spectral signatures of DDoS attacks that can be used to detect repeated instances of such attacks. These signatures have been validated with real attack traces. Specific accomplishments:
  - We have defined methodology to create attack signatures
  - The signatures have been validated with real attack traces

  - We carried sensitivity analysis varying the underlying OS, packet size and number of attackers

The following tasks planned in the proposal have been addressed during this Phase 1 effort:

Task 1: ***Development of an efficient Bayesian approach for recognition of patterns and trends***

Task 2a: ***Motion detection from a moving platform***

Task 2b: ***Multiple target tracking***

Task 3a: ***Developing prototype scenarios for a moderate number of agents in the Hats domain***

Task 3b: ***Building simple probabilistic models for tracking plans and intentions in those scenarios***

Task 3c: ***Improving the Hats Simulator to handle scenarios with as many as*** 10^6 ***agents***

Task 4a: ***Development of new spatio-temporal segmentation and video tracking algorithms***

Task 5a: ***Development of optimal adaptive parametric and nonparametric change-point detection (CPD) procedures***

Task 5b: ***Development of CPD algorithms for the general pattern change process***

Task 5c: ***Energy efficient sensing (detection and tracking)***

Task 6a: ***Real-time integration of information from the variety of surveillance and sensor platforms***

Task 6b: ***Algorithmic development for agile sensors***

Task 7a: ***Attack signature definition***

Task 7b: ***Attack signature validation with real-life attacks***

Task 7c: ***Implement advanced statistical methods such as NLF, stochastic data fusion, and sequential change-point detection to design a scalable forensic IDS with improved capabilities in ultrahigh speed networks***

3. RESEARCH SIGNIFICANCE AND SCIENTIFIC BARRIERS

3.1. Significance

3.1.1. Spatio-Temporal Nonlinear Filtering. So far NLF /HMM techniques have focused mostly on state variable of small dimensionality with point-wise measurements, and henceforth discarded the spatial component of the information while focusing on its dynamic. We plan developing and testing new nonlinear techniques that rely on both spatial and temporal properties. Development of this technology will provide tools that can analyze simultaneously large number of targets or complicated evolving spatial patterns based on incomplete noisy observations. In particular we will derive Zakai and Kushner equations for spatio-temporal random fields and develop numerical methods for their solution.

The main difficulty in practical implementation of spatio-temporal NLF is the computational complexity. It grows dramatically with the introduction of spatial component. We will rely on sequential Markov Chain Monte Carlo (MCMC) methodology (for particle filters) complimented by interacting multiple models for the state dynamics. The latter technique allows for very substantial reduction of the (effective) dimension. To accelerate the MC we will use preliminary filtering based on the Wiener Chaos Expansion. This approach allows one to speed-up the algorithm by shifting off line the time consuming operations related to the prediction steps. In addition, we will develop versions of Kushner and Zakai equations for infinite-dimensional systems.

3.1.2. A Multitarget Tracking Concept for Distributed Targets With Unknown Shapes. Multiple target detection and tracking is a fundamental problem in video surveillance, as it provides the description of spatio-temporal relationships among moving objects in the scene. This information is acquired by many other surveillance modules, e.g., activity recognition.

3.1.3. Tracking Algorithms in Symbolic Spaces and Hats Simulator. The ability to automatically monitor and infer adversary plans/intentions is of great importance for many applications related to the national security. Existing approaches are severely limited in scale (e.g., tracking single agent plans), and use overly simplified models of hostile agent behavior (e.g., no active deception). Our research aims to overcome those limitations, and develop scalable algorithms that will be of practical use to intelligence analysis community.

3.1.4. Pattern Change and Trend Detection in Distributed Multisensor Systems With Applications to Network Security and Surveillance. Decentralized decision making problems are known to be difficult. Without certain conditional independence assumptions across sensors, the problem of finding the optimal solutions, even in the simplest case of static binary hypothesis testing, is computationally intractable. Decentralized dynamic decision making problems, of which the change detection problem is a special case, are even more challenging since they fall into the class of "Witsenhausen problems" with non-classical information patterns. Pattern change and trend detection in distributed sensor networks requires a non-trivial extension of optimal hypothesis testing, change detection, and nonlinear filtering to distributed decentralized systems/scenarios, and implementation to IA and surveillance.

3.1.5. Energy-Efficient Tracking in Sensor Networks. Advances in technology are enabling the deployment of vast sensor networks through the mass production of cheap wireless sensor units with small batteries. Such sensor networks can be used in a variety of application areas. Our focus in this part of the project is on applications of sensor networks that involve *tracking*. The sensor nodes typically need to operate on limited energy budgets. In order to conserve energy, the sensors may be put into a sleep mode. It is clear that the performance of the sensor network could degrade due to having sleepy sensors and therefore any sleeping policy trades off performance with energy savings. Such sleeping is usually effective only if the sensor is completely turned off in the sleep mode, i.e., a sensor that is asleep cannot be communicated with or woken up prematurely. A natural way to implement the sleeping in this setting is to have the sensor enter and exit the sleep mode using a fixed or random duty cycle. We are pursuing an alternative smart approach to sleeping that uses all available information about the state of the network to set the sleep times of the sensors. We have shown through some simple tracking examples that the smart approach can yield significant improvements over the duty cycle approach in the tradeoff between performance and energy savings.

3.1.6. Information Integration and Fusion in Distributed Heterogeneous Multisource Multisensor Systems. The research has significance to design of mobile sensor networks, design of video tracking algorithms for objects with occlusions, and spatio-temporal crime patterns with potential application to terrorist cells.

3.1.7. Intrusion Detection and Networking Problems in IA. Intrusion detection and network attacks form a cat and mouse game where attackers will always devise new ways to evade current defenses. It is thus important to try to stay ahead of the curve. While the number of encrypted, low-level and proxy attacks seen in the wild is not yet large, indications are that they will get larger soon. For example, low-level attacks are easily done today because the average size of a botnet has increased, with some reaching in the millions; several P2P applications have already started using encryption to evade rate limiting; and proxies are prevalent in the Internet today. Clearly, we need to investigate new methods of creating attack signatures that are robust to emerging attacks. Spectral signatures are very promising in this direction. Furthermore, a hybrid approach that combines spectral signatures and statistical change detection in one unit is extremely promising, since this approach allows not only for rapid attach detection but also for an additional false alarm filtering.

3.2. Scientific Barriers

3.2.1. Spatio-Temporal Nonlinear Filtering. Current methodology of pattern recognition in monitoring and surveillance, including network monitoring, geared towards recognizing stationary patterns is not well adapted for coping with detecting emerging patterns in ever-changing environment. The patterns of interest are often spatio-temporal and non-stationary, requiring the development of new methodology allowing learning new trends, as well as recognizing unusual patterns of activity from a heterogeneous data set.

3.2.2. Tracking Distributed Targets With Unknown Shapes. Most existing multiple target tracking methods consider a one-to-one mapping between targets and detected regions, which assume that at a given time instant one observation can be associated with at most one target and vice versa: one target correspond to at most one observation. This assumption is reasonable when the considered observations are punctual, however in video tracking problem, the observations correspond to blobs or meaningful regions which cannot be modeled faithfully by a single point. Moreover, erroneous detections due to occlusion, spurious motion segmentation, or parallax, provide a set of observations where often a single moving object is detected as multiple moving regions, or multiple moving regions are merged into a single blob. The one-to-one association is usually violated in real environments. The spatio-temporal segmentation of objects trajectories relies on the aggregation of hypothesis in time and space for inferring the path of each moving object in the scene. The numerical complexity of the association scheme is therefore substantially large. To solve this combinatorial optimization problem, a Markov Chain Monte Carlo (MCMC) method is proposed to sample the solution space.

3.2.3. Tracking in Symbolic Spaces and Hats Simulator. There are three challenges that we address below.

A. Tracking Collaborative Plans. Most of the existing work on plan-recognition deals with single agent scenarios. In real world situation, malicious plans, such as terrorist attacks, are often executed by teams of hostile agents. Hence, successful algorithms have to take into account possible correlations between actions of various agents, and hypothesize about possible teams of agents that might be involved in hostile activities. This makes the detection problem much more complex.

B. Modeling Deceptive Behavior. Deception is a crucial ingredient of any covert activity. An important (and required) element of a deceptive behavior is to hide certain actions from an observer. There are, however, much more sophisticated forms of deception (e.g., fake build-ups and numerous "invasions" by the allied forces that proceeded the D-day). Existing research in plan recognition has not adequately addressed this type of active deception. To account for this, we will need to develop a formal theory of deception that will allow, among other things, to characterize various forms of deception, and provide computational models of deceptive behavior.

C. Scalability. Even in single agent plan recognition, the complexity of tracking algorithms can be an important issue if the covert agents are embedded in a large benign population. Considering collaborative plans adds another dimension to the problem complexity, as the number of hypothesis about possible covert taskforce grows exponentially with team size.

3.2.4. Change and Trend Detection in Distributed Multisensor Networks. Distributed decentralized decision-making problems (hypothesis testing, estimation, and joint testing-estimation) are extremely non-trivial even when centralized solutions are available. Optimal solutions are barely tractable, and the major goal here is obtaining asymptotically optimal or suboptimal solutions, e.g., a globally asymptotically optimal solution to a change-point detection problem, which is a dynamic decision making problem.

3.2.5. Designing Energy-Efficient Tracking Policies. Designing sleeping policies for energy-efficient tracking involves solving a stochastic control problem (on a potentially large state space) jointly with the nonlinear filtering problem that naturally arises in tracking. The state-space for the control problem grows exponentially with the number of sensors, and so optimal approaches via dynamic programming (DP) are intractable for more than a few sensors. Fortunately we have been able to design provably good suboptimal policies with linear complexity in the number of sensors for a simple sensing and object movement model. There are several challenges that remain to be addressed including: more realistic sensing models, more realistic object movement models, partially known or unknown statistics for object movement, decentralized implementation across sensors, and tracking multiple objects simultaneously.

3.2.6. Information Integration in Distributed Heterogeneous Multisource Multisensor Systems. The challenges faced by the team included the hardware design of a platform with micro-sized vehicles to process data and information in real time. Also we considered video footage in which an object of interest goes behind an occlusion and must continue to be tracked. Finally we built models for crime data from scratch based on individual movement of criminals and likelihood of them breaking into a location due to previous history of that location.

3.2.7. Intrusion Detection and Networking Problems in IA. Three aspects have to be emphasized:

A. Real Data Collection – LANDER. Up to this point there were no real attack data sets available to test and validate our detection algorithms. Recently this has changed with the creation of LANDER (Los Angeles Network Data Exchange and Repository), a DHS-funded effort to collect and distribute real attack data sets. LANDER is based at USC and we have full access to all data in the repository. In addition to collecting data from the regional ISP serving USC and other educational and commercial institutions, LANDER has recently installed a capture point at a commercial web-hosting service that routinely gets attacked on a daily basis. We expect that with LANDER we will solve the problem of access to real-world attack data.

B. Real-time Spectral Signature Generation. The other challenge is adapting standard signal processing techniques to time series generated from network traffic. The challenges there include using the right variables to track in a time series, the right resolution, etc.

C. Hybrid Anomaly-Signature IDS. False alarm rate (FAR) of anomaly-based detectors with hard decisions may be improved by analyzing more detailed patterns in traffic statistics, i.e., signatures. Therefore, combining spectral signature approach and corresponding signal processing techniques with anomaly change detection based techniques seems to be beneficial. This approach is complementary to the anomaly-based and signature-based IDSs and allows for profiling, i.e., confirmation or rejection of detection decisions at the output of the anomaly detector using signature analysis. Combining these two methods into a hybrid IDS is not a trivial task.

4. TECHNICAL APPROACH AND MAIN RESULTS OF THE PHASE 1 EFFORT

4.1. Efficient Bayesian Approach for Recognition of Patterns and Trends — Spatio-Temporal Nonlinear Filtering

The analysis and synthesis of high-volume data and in particular spatio-temporal data is addressed by different disciplines and from different perspectives. Our approach is probabilistic in nature, more specifically it is Bayesian. One important feature of the Bayesian approach is that it interprets the data not as a self-contained information depository but in the light of already available knowledge (e.g., human expertise) regarding the events reflected by the data. This feature is an ideal instrument for keeping human operators “in the loop” in the process of automated decision making.

Nonlinear filtering (NLF) is an extension of the Bayesian framework into dynamical systems. Kalman filter designed for linear dynamical systems and linearly structured observations is probably the most famous Bayesian filter. Its generalizations to nonlinear systems and/or observations is usually referred to as optimal nonlinear filtering (ONLF) for hidden Markov models (HMM). ONLF is a field on the cutting edge of contemporary stochastic analysis, information theory, and statistical inference. This is an emerging methodology with enormous breadth of applications.

So far NLF/HMM applications have focused mostly on state processes of small to medium complexity insufficient for the applications of interest for this project. One of the main objectives of this project is development of spatio-temporal NLF algorithms and their DoD relevant applications. Certain progress in this direction has been made during the current stage of the grant. This progress is outlined below.

To address tracking objects in distributed data/images we have derived a preliminary version of Zakai and Kushner equations for spatial-temporal observation process with continuous and discrete observations. In particular, we have derived the analogs of Zakai and Kushner equations of nonlinear filtering in this setting.

Applications to tracking multiple agents/plans (see Section 4.3) requires development of nonlinear filtering methods for telescoping Markov processes. The simplest example of this type of processes is referred to as interacting multiple models (IMM). Our recent results allow for extending this methodology to very complicated systems. Partial testing of the obtained algorithm was performed on simple "Hats" models.

One of the very difficult problems of nonlinear filtering is related to estimation of processes related to uncertainty of the observation. This is often modeled as a process controlling the intensity of the noise in observations. We have proposed and developed an NLF algorithm for this setting with observations obtained at random times. The results turned out to be very promising. They were applied to tracking volatility (see papers [9] and [10]). A similar algorithm for tracking the level of hostile "chatter" on the Internet is under consideration.

4.2. Video Tracking Multiple Distributed Targets From a Moving Platform

The overview framework of detection and tracking multiple target from a moving platform is shown in Figure 1.

4.2.1. Image Registration and Adaptive Background Modeling. The main difference between the detection of moving objects from a stationary and moving camera is the characterization of the background model. In a stationary camera, variations in the image sequence are modeled at the pixel level and allow defining a background model for each pixel using statistical-based techniques. This concept can be extended to non-stationary cameras by compensating for the camera motion prior to the estimation of the background model. Registering the current frame to the selected reference is performed by concatenating the estimated pair-wise transforms, shown in Figure 2. We propose to establish the background modeling within a sliding window to reduce the accumulated registration error.

4.2.2. Boosted MCMC Data Association for Multiple Target Tracking. We formulate the multiple targets tracking as an association problem, in which the purpose is to find the best spatio-temporal association (shown in Figure 3) between observations (i.e., detected moving regions) and targets while maximizing the posterior association probability. This spatio-temporal association method which does not require the one-to-one mapping between observations and targets. We represent the association problem in a deferred logic way where association is defined between targets and a set of latest observations within a sliding window. This allows the association decision to be made when enough observations are available. As the size of sliding window grows, the scale of the problem grows exponentially. To avoid the enumeration of all possible association hypothesis and to solve this combinatorial optimization problem efficiently, we propose an Markov Chain Monte Carlo (MCMC) [13] method to sample the solution space. Instead of separating the detection and tracking as two separate procedures, each preliminary detection derived by the motion segmentation is assigned a model likelihood provided by a real-valued Adaboost classifier trained offline. The MCMC sampling is driven by an informed proposal scheme controlled by a joint

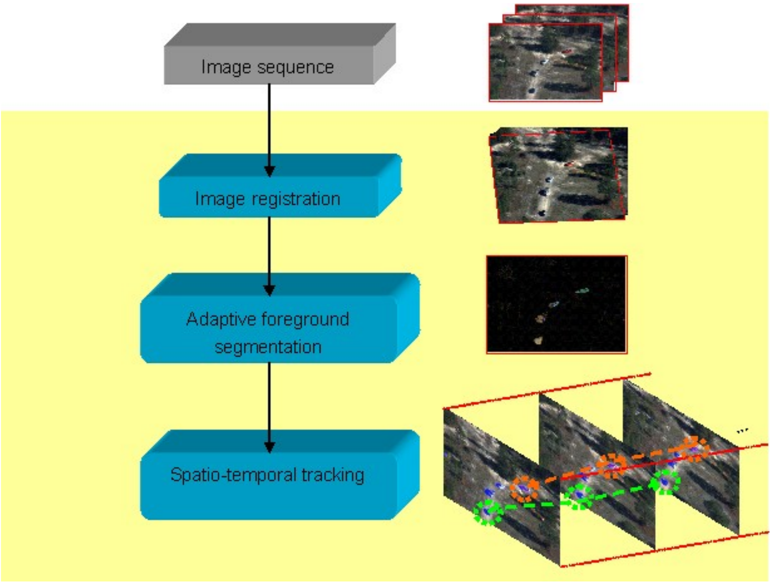


Figure 1: The framework of multiple target detection and tracking from a moving platform

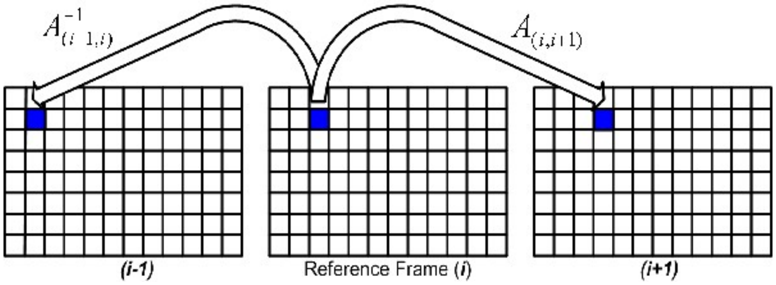


Figure 2: Adaptive background modeling

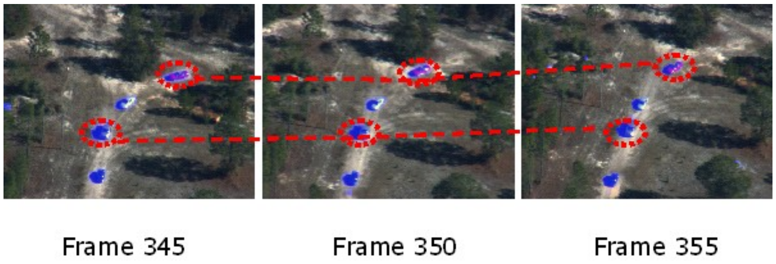


Figure 3: Spatio-temporal association tracking

probability model combining motion, appearance and model likelihood among detected regions. We test our detection and tracking framework on videos captured by moving platform such as Unmanned Aerial Vehicles (UAV).

4.3. Probabilistic Tracking and Detecting Hostile activities in Hats

Our probabilistic framework will be designed to detect and track hostile plans and intentions in a virtual society of agents. We assume that most of the agents are engaged in benign activities, while a small number have malicious intent. Some specific research questions in this regard are:

- How does the detection accuracy depend on the difference between benign and covert behavior?
- How much data does one need for detecting malicious intent?
- What kind of error rates should one expect?

The main goal of the proposed research is to obtain quantitative answers to those questions, and use the insights from the analysis to develop efficient and scalable probabilistic algorithms for detection and tracking in symbolic space of plans and intentions.

4.3.1. The Probabilistic Framework: Model and Inference. The theoretical foundation of our tracking model is provided by an *Abstract Hidden Markov Model* (AHMM) [5]) that uses a Dynamic Bayesian Network representation of the plan hierarchy. There is a set of possible states *S* which is called the state space. At each state *s*, an agent has a set of actions *A* available, where each action *a*, if employed, will cause the system to move to the next state s' via a transition probability $\sigma_a(s, s')$. An agent's plan of action is modeled as a *policy* that determines which action the agent will choose at each state. For a policy π, this is modeled by a selection function $\sigma_\pi : S \times A \to [0, 1]$, where at each state s, $\sigma_\pi(s, a)$ is the probability that the agent will choose the action a. Thus, for a fixed policy π, the resulting state sequence is a Markov chain with transition probabilities $Pr(s'|s) = \sum_a \sigma_\pi(s, a)\sigma_a(s, s')$. Hence a policy can also be viewed as a Markov chain through the state space. A policy hierarchy is defined as a sequence $H = (\Pi_0, \Pi_1, \ldots, \Pi_K)$ where K is the number of levels in the hierarchy, Π_0 is a set of primitive actions, and for $k = 1, \ldots, K$, Π_k is a set of policies over the policies in Π_{k-1}. When a top-level policy π^K is executed, it invokes a sequence of level-$(K-1)$ policies, each of which invokes a sequence of level-$(K-2)$ policies, and so on. A level-1 policy will invoke a sequence of primitive actions which leads to a sequence of states.

We will then adopt the multi-agent extension of this method devised by [22] known as *Hierarchical Multiagent Markov Processes* (HMMP), which assumes that the agents coordinate their actions at more abstract levels by explicitly using a central controller, but at lower levels, individual policies are executed without coordination by each agent. This will help us build a framework for tracking subsets of agents together, thus leading to the detection of potential task forces for a harmful event.

In the framework of AHMM, it is assumed that a policy hierarchy is given, however, the top level policy and the details of its execution are unknown. The problem is to determine the top level policy and other current policies at the lower levels given the current sequence of observations. In other words, we are interested in estimating the conditional probability $\Pr(\pi_t^K, \ldots, \pi_t^0|\tilde{o}_{t-1})$, and the marginals $\Pr(\pi_t^k|\tilde{o}_{t-1})$, for all levels k. This is done by updating the belief state $\Pr(\pi_{t+1}^k, o_{t+1}|\tilde{o}_t)$

at each state using the posterior after absorbing the observation o_{t+1} at time $t+1$. Computing these probabilities gives us the information about the current policies at all levels of abstraction, from the current action $(k = 0)$ to the top level policy $(k = K)$, taking into account all the observations that we have up to date. Computing these probabilities is generally intractable unless the belief state has an efficient representation that affords a closed-form update procedure. Without any structure imposed on the belief state, the complexity for updating it is exponential in K. To cope with this complexity, a hybrid inference scheme based on *Rao-Blacwellized Sequential Importance Sampling* (RB-SIS; [5]) is used, which combines both approximation and tractable exact inference for efficiency.

4.3.2. The Hats Domain. The Hats Simulator [8] is designed to be a lightweight proxy for many intelligence analysis problems, and thus as a test environment for analysts' tools. It is a virtual world in which millions of agents engage in individual and collective activities. Most are benign and a small fraction of them intend harm, and each hat belongs to one or more organizations (benign and terrorist). The activities are planned by a generative planner and the job of the analyst is to find harmful agents before they can attack certain landmarks referred to as *beacons*. The beacons have known vulnerabilities that must be acquired by the task force (TF) in order to be able to destroy it. First, the planner chooses a target beacon for the attack, followed by the choice of a TF, and then assigns roles to TF members, i.e., each member is assigned a certain capability that will be transported to the beacon at the final stage of the attack. Once these assignments are made, the planner generates a meeting schedule for each TF member, so that on completion of these meetings, each TF carries the assigned capability. Finally, the TF is moved to the beacon location for the final meeting or the attack. A simple Hierarchical Task Network (HTN) representing the HATS scenario is depicted in Figure 4. Note that, during plan execution, only the last step produces actual observations (highlighted nodes in Figure 4) in the form of meetings, while the choice

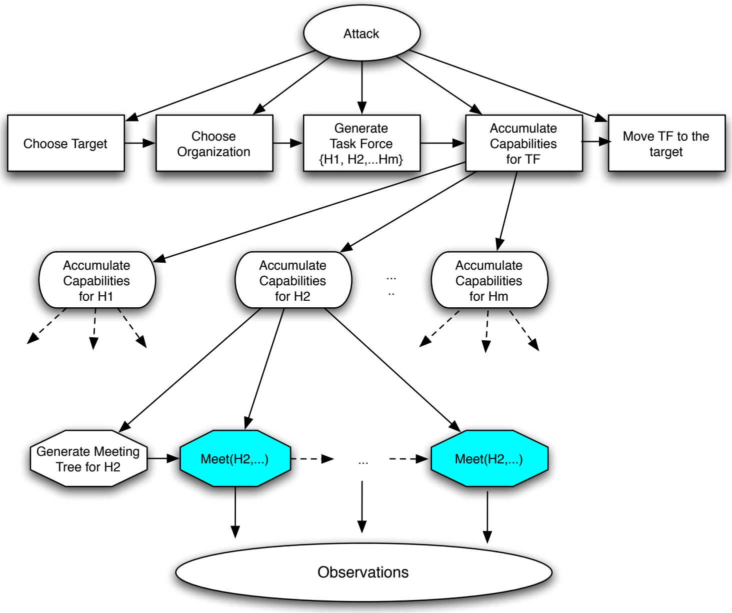

Attack
Choose Target
Choose Organization
Generate Task Force {H1, H2,...Hm}
Accumulate Capabilities for TF
Move TF to the target
Accumulate Capabilities for H1
Accumulate Capabilities for H2
Accumulate Capabilities for Hm
Generate Meeting Tree for H2
Meet(H2,...)
...
Meet(H2,...)
Observations

Figure 4: The HTN representation of a terrorist attack planning in HATS

of the target, the TF, role assignment and the capability trades are all hidden from the analyst. However, the simulator maintains information about all the agents that may be acquired through an *information broker* by paying a certain “fee”.

4.3.3. The AHMM Formulation in HATS. In the HATS domain, each policy in the plan hierarchy is represented by a node in Figure 4. The meetings between the hats are the observable actions used to update the beliefs about whether a particular hat possesses the required capabilities needed to attack the beacon at each time point. So far, we have developed a simple set of update rules based on the meeting trees of each agent that are executed independently (since TF members meet only at the final meeting). The single-agent policies therefore comprise of a hat acquiring a set of capabilities with the trades having a strict temporal ordering constraint that evolves over time. There are 4 levels of hierarchy in the HATS domain ($K = 3$) – (i) level-0, which has the observations (meetings), (ii) level-1, which is the state where each agent is acquiring capabilities, (iii) level-2, where all the required capabilities have been acquired by the TF members, and (iv) level-3, when the final meeting (attack) occurs.

4.3.4. The Bayesian Update Rules. Let us consider a population of hats $\{H_1, H_2, \ldots\}$, and let $\{C_1, C_2, \ldots, C_N\}$ denote the complete set of N capabilities that these hats may possess. Let us assume for the time being that the analyst has some prior beliefs about the initial capabilities that each H_i has. The hats then go to meetings $\{M_i^1, \ldots, M_i^{T_i}\}$ with other hats. Since only meetings are observed and not the trades, we can only assume that a trade occurs with a certain probability at each time between two hats. More precisely, $p^j_{\{i,k\}} = \Pr(H_k \text{ trades } C_j \text{ to } H_i | H_k \text{ has } C_j)$, if H_i meets with H_k at a time point. For now, we assume this to be constant, that is, $p^j_{\{i,k\}} = p_{tr}$ for all i, j, k. Now let us define a random variable $C^i_{j,t}$ such that

$$C^i_{j,t} = \begin{cases} 1, & \text{if } H_i \text{ has } C_j \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

We then wish to update the belief of this event at time t given by $B_t^{i,j} = Pr(C_{j,t} = 1 | \tilde{M}_i^t)$, where $\tilde{M}_i^t$ denotes all the meetings of hat H_i till time t, thus giving rise to the posterior after absorbing the observation (meeting with H_k) at the next time point $(t+1)$ M_i^{t+1}:

$$\begin{aligned} B^{i,j}_{t+1} &= Pr(C^i_{j,t+1} = 1 | \tilde{M}^i_{t+1}) = Pr(C^i_{j,t} = 1 | \tilde{M}^t_i) \times Pr(H_i \text{ does not trade } C_j \text{ to } H_k | C^i_{j,t} = 1) \\ &+ Pr(C^i_{j,t} = 0 | \tilde{M}^i_t) \times Pr(H_i \text{ acquires } C_j \text{ at time } t+1 | \tilde{M}^i_{t+1}) \\ &= B^{i,j}_t (1 - p_{tr}) + [1 - B^{i,j}_t] B^{k,j}_t p_{tr}. \end{aligned} \tag{2}$$

Now, in case H_i does not participate in a trade at time $(t+1)$, there is no update to the belief about its capabilities at that time point, so that $B^{i,j}_{t+1} = B^{i,j}_t$, for all j. Hence we have a nice general rule for updating the belief at any time t recursively for each hat and each capability individually.

However, in the hats scenario, the capabilities of a hat are subject to decay or expiration and in order to incorporate this in Equation 2, we define $\delta^{i,j}_t$ to be the probability of decay of capability C_j for H_i at time point t. Then the revised update rule for $B^{i,j}_t$ on absorbing the observation at time $(t+1)$ can be written as

$$\begin{aligned} B^{i,j}_{t+1} &= B^{i,j}_t (1 - p^j_{\{k,i\}}) + [1 - B^{i,j}_t] B^{k,j}_t p^j_{\{i,k\}} + Pr(C_{j,t} = 1 | \tilde{M}^i_t) \\ &\times Pr(C_j \text{ does not expire for } H_i | C^i_{j,t} = 1) = B^{i,j}_t (2 - p_{tr} - \delta^{i,j}_t) + [1 - B^{i,j}_t] B^{k,j}_t p_{tr}. \end{aligned}$$

In case H_i does not participate in a trade at time $(t+1)$, the revised update rule now will be,

$$B^{i,j}_{t+1} \;=\; Pr(H_i \text{ has } C_j \text{ at time } t \text{ and it does not expire}|\tilde{M}^i_{t+1}) \;=\; B^{i,j}_t(1-\delta^{i,j}_t).$$

Let us now consider a situation with M beacons and P terrorist organizations. The next task therefore is to propagate the beliefs about hats possessing different capabilities to determine how likely a hat is to plan an attack on a particular beacon and be a member of a particular terrorist organization. The probability of H_i attacking Beacon r with a vulnerability set L can be computed as:

$$\begin{aligned} P^{i,r}_{t+1} &= Pr(H_i \text{ has at least 1 of the vulnerabilities at time } t+1|\tilde{M}^i_{t+1}) \\ &= 1 - Pr(H_i \text{ has none of the 3 required capabilities}|\tilde{M}^i_{t+1}) \\ &= 1 - \Pi_{l\in L}(1-B^{i,l}_{t+1}). \end{aligned} \tag{3}$$

Note that, we have made the approximation here that the probability of a TF member possessing none of the vulnerabilities of the beacon is negligible, and assumed that hats possess capabilities independently. Figure 5 shows a simple schematic how the new evidence based on meetings are gradually propagated up the plan hierarchy through the Bayesian updating-filtering mechanism that we just described.

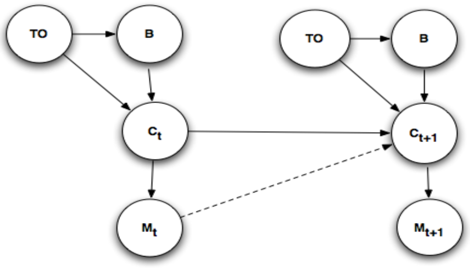


Figure 5: The conditional independence relationships in the HATS domain. “TO” denotes the terrorist organization, “B” denotes the beacon to attack, “c_t” denotes the capabilities and “M_t” denotes the meetings.

4.3.5. Future work. In our proposed research we will refine this simple model to achieve a more general and coherent framework for detection and tracking. In particular, we will introduce joint update rules for all capabilities, intentions to attack a particular beacon, as well as terrorist/benign status of an agent. This is expected to enhance the accuracy of our proposed model by exploiting the conditional independence relationships. We will compute the joint likelihood

$$Pr(\pi^K_t, \ldots, \pi^0_t|\tilde{o}_{t-1})$$

using the AHMM framework and perform inference relating to the hypotheses about the potential task forces. Note that, the update rules have closed-form expressions here and hence tractable inference will be possible. Another aspect of the research will involve incorporating a “hidden” variable to denote the intention of each hat to acquire the required attributes of a beacon. Three other significant research directions involve:

(i) identify deceptive behavior by developing a rigorous theoretical framework of deception detection,

(ii) introduce coordination among the members of the task force in a multi-agent scenario using the HMMP model, and

(iii) scalability and other computational issues that will assess how well our model treats a large number of hats in the order of millions.

4.4. Pattern Change and Trend Detection in Distributed Sensor Networks

One of the goals of this project is to develop new procedures for pattern change and trend detection in distributed multisensor systems, and to provide an analytical framework to predict their performance in terms of the tradeoff between detection delay and frequency of false alarms.

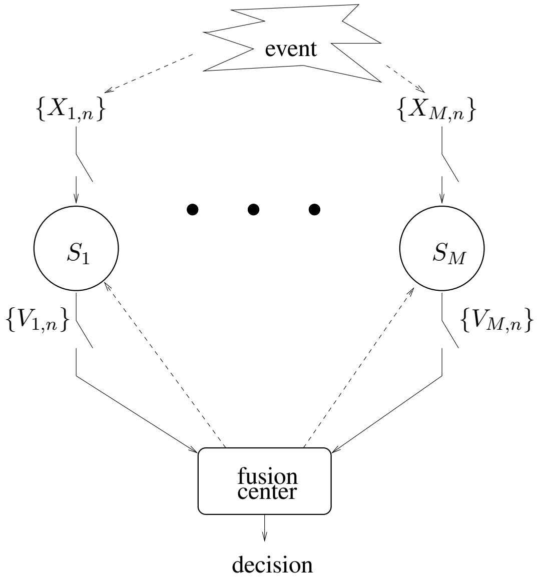


Figure 6: Change detection using distributed sensors and modes of operation

To address this goal, we performed analysis of several generalizations of the change detection problem that arise in the applications to distributed sensor systems. In the distributed multisensor systems that are the focus of this project, the information about the change is available through a set of geographically separated sensors, as shown in Figure 6. Specifically, we consider the distributed multisensor system with N sensors, $S_1, \dots, S_N$, communicating with a fusion center. At time n, an observation $X_{i,n}$ is made at sensor S_i. The changes in the statistical properties of the sequences $\{X_{i,n}\}$ are governed by the event. The sensors communicate to a common fusion center. We investigate a variety of models for the change process: only one (or a subset) of the sensors changes, they all change at the same time, or they change at different times. We also include various scenarios for communication with the fusion center, from the centralized one where the sensors send sufficient statistics, to the decentralized one where they send quantized observations or local decisions. We study the role of feedback from the fusion center, and investigate schemes for conserving energy at the sensors such as switching the sensors between on/off modes and censoring their observations. Our strategy for design and analysis accommodate general statistical models for the observations, and allow for different degrees of model uncertainty.

Based on the information available at S_i at time n, a message $V_{i,n}$ is sent from sensor S_i to the fusion center. There are various possibilities for communication from the sensors: they could send the observations (or sufficient statistics), they could send quantized observations, and they could choose to censor their observations. Censoring refers to the situation where the sensors refrain from sending any information to the fusion center for certain observations; this is indicated by switches on $V_{i,n}$ in Figure 6. Also, to conserve energy the sensors may switch between on and off modes; this is indicated by switches on $X_{i,n}$.

This concert of possibilities leads to a very interesting set of open problems that will be discussed in the course of future research. In order to address the wide range of potential applications of our theory, we will accommodate general statistical models for the observations and allow for different degrees of model uncertainty.

In the rest of this section, we will be interested in a particular distributed and decentralized multisensor scenario where no communication between sensors and no feedback between the fusion center and sensors are allowed, as shown in Figure 7. The statistical properties of the sensors' observations change at the same unknown point in time. The goal is to detect this change as soon as possible, subject to false alarm constraints. The sensors may send either quantized versions of their observations or local decisions to a fusion center where a final decision is made based on all the sensor messages.

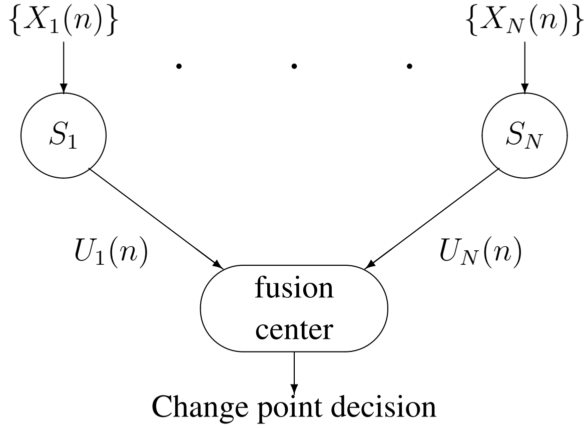


Figure 7: Change detection with distributed sensors and no feedback

Therefore, there is a distributed N-sensor system in which at time n one observes an N-component vector stochastic process $(X_1(n), \ldots, X_N(n))$. The i-th component $X_i(n)$, $n = 1, 2, \ldots$ corresponds to observations obtained from the sensor S_i, as shown in Figure 7. We will consider two approaches to the decentralized fusion problem. In the first case, the sensors quantize their observations and these quantized observations are sent to the fusion center; in the second scenario they make local decisions that are sent to the fusion center. At an unknown point in time λ ($\lambda = 1, 2 \ldots$) something happens and all of the components change their distribution. Conditioned on the change point, the observation sequences $\{X_1(n)\}$, $\{X_2(n)\}$, $\ldots$, $\{X_N(n)\}$ are assumed to be mutually independent. Moreover, we assume that, in a particular sensor, the observations are independent and identically distributed (iid) before and after the change (with different distributions). If the change occurs at $\lambda = k$, then in sensor S_i the data $X_i(1), \dots, X_i(k-1)$ follow the distribution F_0^i with a density $f_0^{(i)}(x)$, while the data $X_i(k), X_i(k+1), \dots$ have the common distribution F_1^i with a density $f_1^{(i)}(x)$ (both with respect to a sigma-finite measure $\mu(x)$).

To be more specific, let $\mathbf{P}_k$ (correspondingly $\mathbf{E}_k$) be the probability measure (correspondingly expectation) when the change occurs at time $\lambda = k$. Then, $\mathbf{P}_\infty$ and $\mathbf{E}_\infty$ stand for the probability measure and expectation when $\lambda = \infty$, i.e., the change does not occur. Write $\mathbf{X}_i^n = (X_i(1), \dots, X_i(n))$ and $\mathbf{X}^n = (\mathbf{X}_1^n, \dots, \mathbf{X}_N^n)$. Under $\mathbf{P}_\infty$, the density of $\mathbf{X}^n$ is

$$p_0(\mathbf{X}^n) = \prod_{i=1}^{N} \prod_{j=1}^{n} f_0^{(i)}(X_i(j)) \text{ for all } n \geqslant 1$$

and, under $\mathbf{P}_k$, the density of $\mathbf{X}^n$ is

$$p_k(\mathbf{X}^n) = \prod_{i=1}^{N} \left[\prod_{j=1}^{k-1} f_0^{(i)}(X_i(j)) \prod_{j=k}^{n} f_1^{(i)}(X_i(j)) \right]$$

for $k \leqslant n$ and $p_k(\mathbf{X}^n) = p_0(\mathbf{X}^n)$ for $k > n$.

In the minimax setting, a reasonable measure of the detection lag is the supremum average detection delay (SADD)

$$\mathrm{SADD}(\tau) = \sup_{1 \leqslant k < \infty} \mathbf{E}_k(\tau - k | \tau \geqslant k)$$

while the false alarm rate can be measured by the average run length (ARL) to false alarm

$$\mathrm{ARL}(\tau) = \mathbf{E}_\infty \tau.$$

An optimal minimax detection procedure is a procedure for which $\mathrm{SADD}(\tau)$ is minimized while $\mathrm{ARL}(\tau)$ is set at a given level γ, $\gamma > 0$. Specifically, define the class of change-point detection procedures

$$\boldsymbol{\Delta}(\gamma) = \{\tau : \mathrm{ARL}(\tau) \geqslant \gamma\}$$

for which the ARL exceeds the predefined positive number γ. The optimal change-point detection procedure is described by the stopping time

$$\nu = \arg \inf_{\tau \in \boldsymbol{\Delta}(\gamma)} \mathrm{SADD}(\tau).$$

Let

$$Z_i(n) = \log \frac{f_1^{(i)}(X_i(n))}{f_0^{(i)}(X_i(n))} \tag{4}$$

be the log-likelihood ratio (LLR) between the “change” and “no-change” hypotheses for the n-th observation from the i-th sensor and let

$$\mathrm{I}_i = \mathbf{E}_1 Z_i(1) = \int \log \left(\frac{f_1^{(i)}(x)}{f_0^{(i)}(x)} \right) f_1^{(i)}(x) \mu(dx)$$

be the Kullback-Leibler (K-L) information number between the densities $f_1^{(i)}(x)$ and $f_0^{(i)}(x)$.

The asymptotic performance of an optimal centralized detection procedure that has access to all data $\mathbf{X}^n$ is given by

$$\inf_{\tau \in \boldsymbol{\Delta}(\gamma)} \mathrm{SADD}(\tau) = \frac{\log \gamma}{\mathrm{I}_{\mathrm{tot}}}(1 + o(1)), \quad \gamma \to \infty, \tag{5}$$

where $\mathrm{I}_{\mathrm{tot}} = \sum_{i=1}^N \mathrm{I}_i$. See, e.g., [2, 23, 24]. This performance is attained for the centralized CUSUM test that uses all available data.

4.4.1. Centralized CUSUM Detection Test. The centralized CUSUM test is defined as

$$\tau_c = \min \left\{ n \geqslant 1 : W^c(n) \geqslant h \right\},$$

where the (centralized) CUSUM statistic $W^c(n)$ is given by the recursion

$$W^c(n) = \max \left\{ 0, W^c(n-1) + \sum_{i=1}^N Z_i(n) \right\}, \quad n = 1, 2, \dots \tag{6}$$

($W^c(0) = 0$) and the threshold h is chosen so that $\mathrm{ARL}(\tau_c(h)) = \gamma$. It is known [2, 23, 24] that $\mathrm{ARL}(\tau_c(h)) \geqslant e^h$ and, hence, $h = \log \gamma$ guarantees $\mathrm{ARL}(\tau_c(h)) \geqslant \gamma$. The latter choice is usually conservative but useful for preliminary estimates and first-order asymptotic analysis. Substantial improvements can be obtained using corrected Brownian motion approximations [23] and the renewal argument [25].

In the following two subsections, we consider two types of decentralized detection procedures that use "compressed" data $U_1(n), \dots, U_N(n)$ which are transmitted to the fusion center for making the final decision. The compression level for both types of procedures is maximal – the data $U_i(n) = 0$ or 1, i.e., binary. Thus, for both proposed decentralized detection procedures the required bandwidth for communication with the fusion center is minimal. The advantage of the first detection test with binary quantized data is that it does not require any processing power at the sensors. In Section 4.4.3, even simpler "one-shot voting" local decision (LD) based detection tests are introduced.

4.4.2. Decentralized CUSUM Test with Binary Quantization at the Sensors. Consider the scenario where based on the observation $X_i(n)$ available at the sensor S_i at time n a message $U_i(n)$ belonging to a finite alphabet (e.g., binary) is formed and sent to the fusion center (see Figure 7). Write $\mathbf{U}_n = (U_1(n), \dots, U_N(n))$ for the vector of N messages at time n. Based on the sequence of sensor messages, a decision about the change is made at the fusion center. The goal is to find a detection test at the fusion center that has certain optimality properties. This test is identified with a stopping time on $\{\mathbf{U}_n\}_{n \geqslant 1}$ at which it is declared that a change has occurred.

In the following we consider the simplest case where $U_i(n) = \psi_i(X_i(n))$ are the outputs of binary quantizers. The asymptotically optimal policy for the decentralized change detection problem with binary quantization that minimizes $\mathrm{SADD}(\tau) = \sup_k \mathbf{E}_k\{\tau - k | \tau \geqslant k)$, while maintaining the $\mathrm{ARL}(\tau)$ at a level greater than γ, consists of a set of stationary monotone likelihood ratio quantizers (MLRQ) at the sensors followed by the CUSUM procedure based on $\{\mathbf{U}_n\}_{n \geqslant 1}$ at the fusion center [26, 31].

More specifically, the optimal binary quantizer is the MLRQ which is given by

$$U_i = \psi_i(X) = \begin{cases} 1 & \text{if } \frac{f_1^{(i)}(X)}{f_0^{(i)}(X)} \geqslant t_i, \\ 0 & \text{otherwise,} \end{cases}$$

where t_i is a positive finite threshold that maximizes the K-L information in the resulting Bernoulli sequence for the post-change and pre-change hypotheses.

To be precise, for $l = 0, 1$, let $g_l^{(i)}$ denote the probability induced on $U_i(n)$ when the observation $X_i(n)$ is distributed as $f_l^{(i)}$. Let $\beta_{0,i} = g_0^{(i)}(U_i(j) = 1)$ and $\beta_i = g_1^{(i)}(U_i(j) = 1)$ denote the corresponding probabilities under the normal and the anomalous conditions, respectively. The resulting binary (Bernoulli) sequences $\{U_i(j), i = 1, \dots, N\}$, $j \geqslant 1$ are then used to form the binary CUSUM statistic similar to (6) as

$$W^b(n) = \max\{0, W^b(n-1) + \sum_{i=1}^{N} Z_i^b(n))\}, \quad n = 1, 2, \dots \tag{7}$$

where $W^b(0) = 0$ and

$$Z_i^b(n) = \log \frac{g_1^{(i)}(U_i(n)))}{g_0^{(i)}(U_i(n))}$$

is the partial LLR between the "change" and "no-change" hypotheses for the binary sequence, which is given by

$$Z_i^b(n) = a_i U_i(n) + a_{0,i}.$$

Here

$$a_i = \log \frac{\beta_i(1-\beta_{0,i})}{\beta_{0,i}(1-\beta_i)}, \quad a_{0,i} = \log \frac{1-\beta_i}{1-\beta_{0,i}}.$$

Then the CUSUM detection procedure at the fusion center is given by the stopping time

$$\tau_b(h) = \min\left\{n \geqslant 1 : W^b(n) \geqslant h\right\}, \tag{8}$$

where h is a positive threshold which is selected so that $\mathrm{ARL}(\tau_b(h)) \geqslant \gamma$. In what follows this detection procedure will be referred to as the binary quantized CUSUM test (BQ-CUSUM).

The BQ-CUSUM procedure with $h = \log\gamma$ is asymptotically optimal as $\gamma \to \infty$ in the class of tests with binary quantization in the sense of minimizing the SADD in the class $\boldsymbol{\Delta}(\gamma)$. More specifically, the tradeoff curve for the optimal binary test is

$$\mathrm{SADD}(\tau_b) \sim \frac{\log\gamma}{\mathrm{I}_{\mathrm{tot}}^b}, \quad \gamma \to \infty, \tag{9}$$

where $\mathrm{I}_{\mathrm{tot}}^b = \sum_{i=1}^{N} \max_{t_i} \mathrm{I}_i^b(t_i)$ is the total maximal K-L distance (optimized over the quantization thresholds t_i); $\mathrm{I}_i^b(t_i) = [\beta_i(t_i)a_i(t_i) + a_{0,i}(t_i)]$ is the K-L distance for the binary sequence in the i-th sensor for the quantization threshold t_i.

The asymptotic relative efficiency (ARE) of a detection procedure τ_γ with respect to a detection procedure η_γ, both of which meet the same lower bound γ for the ARL, will be defined as

$$\mathrm{ARE}(\tau_\gamma; \eta_\gamma) = \lim_{\gamma \to \infty} \frac{\mathrm{SADD}(\tau_\gamma)}{\mathrm{SADD}(\eta_\gamma)}.$$

Using (5) and (9), we obtain that the ARE of the globally asymptotically optimal test ν with respect to the BQ-CUSUM test τ_b is

$$\mathrm{ARE}(\nu; \tau_b) = \lim_{\gamma \to \infty} \frac{\inf_{\tau \in \boldsymbol{\Delta}(\gamma)} \mathrm{SADD}(\tau)}{\mathrm{SADD}(\tau_b(h_\gamma))} = \frac{\mathrm{I}_{\mathrm{tot}}^b}{\mathrm{I}_{\mathrm{tot}}}. \tag{10}$$

Since $\mathrm{I}_{\mathrm{tot}}$ is always larger than $\mathrm{I}_{\mathrm{tot}}^b$, the value of $\mathrm{ARE} < 1$. However, our study presented below shows that certain decentralized asymptotically globally optimal tests may perform worse in practically interesting prelimit situations when the false alarm rate is moderately low but not very low.

4.4.3. Decentralized Detection Tests Based on Local Decisions. We now consider three detection schemes that perform local detection in the sensors and then transmit these local binary decisions to the fusion center for optimal combining and final decision-making. The abbreviation LD-CUSUM will be used for procedures that perform CUSUM tests in sensors and use local decisions.

A. Asymptotically globally optimal decentralized LD-CUSUM test. Let

$$W_i(n) = \max\{0, W_i(n-1) + Z_i(n)\}, \quad W_i(0) = 0$$

be the CUSUM statistic in the i-th sensor, where, as before, $Z_i(n) = \log[f_1^{(i)}(X_i(n))/f_0^{(i)}(X_i(n))]$ is the LLR for the original sequence.

Let

$$U_i(n) = \begin{cases} 1 & \text{if } W_n(i) \geqslant \pi_i h \\ 0 & \text{otherwise,} \end{cases}$$

where $\pi_i = \mathrm{I}_i/\mathrm{I}_{\mathrm{tot}}$ and h is a positive threshold.

The stopping time is defined as

$$T_{\mathrm{ld}}(h) = \min\left\{ n : \min_{1 \leqslant i \leqslant N} [W_i(n)/\pi_i] \geqslant h \right\}. \tag{11}$$

In other words, binary local decisions (1 or 0) are transmitted to the fusion center, and the change is declared at the first time when $U_i(n) = 1$ for all sensors $i = 1, \dots, N$.

It can be shown that under certain conditions

$$\mathbf{E}_\infty T_{\mathrm{ld}}(h) \geqslant e^h$$

and

$$\mathrm{SADD}(T_{\mathrm{ld}}(h)) = \frac{h}{I_{\mathrm{tot}}} + C_N \sqrt{\frac{h}{\mathrm{I}_{\mathrm{tot}}}} - 1 + o(1), \tag{12}$$

where

$$C_N = \mathbf{E} \max_{1\leqslant i\leqslant N} \left\{ \frac{\sigma_i}{\mathrm{I}_i} Y_i \right\}, \tag{13}$$

$Y_1, \dots, Y_N$ are independent standard Gaussian random variables; $\sigma_i = \sqrt{\mathrm{Var}_i(Z_1(i))}$; Var_i is the operator of variance under $f_1^{(i)}$.

Therefore, if $h = \log \gamma$, then

$$\inf_{\tau\in\mathbf{\Delta}(\gamma)} \mathrm{SADD}(\tau) \sim \mathrm{SADD}(T_{\mathrm{ld}}(h)) \sim \frac{\log\gamma}{\mathrm{I}_{\mathrm{tot}}}, \quad \gamma\to\infty$$

and the detection test $T_{\mathrm{ld}}(h)$ is globally asymptotically optimal (AO), i.e., $\mathrm{ARE}(T_{\mathrm{ld}}; \tau_c) = 1$. Correspondingly, we will use the abbreviation AO-LD-CUSUM for this test in the rest of the report.

However, since the second term in the asymptotic approximation (12) is on the order of the square root of the threshold, it is expected that the convergence to the optimum is slow. Note that for the optimal centralized CUSUM test and for the decentralized CUSUM test with binary quantization residual terms are constants. We therefore expect that for moderate false alarm rates typical for practical applications the procedure with quantization may perform better. Below this conjecture is verified for the Poisson model.

B. "One-shot" voting decentralized LD-CUSUM tests. Let $\tau_i(h) = \min\{n : W_i(n) \geqslant h\}$ denote the stopping time of the CUSUM test in the i-th sensor. Introduce the stopping times

$$T_{\min}(h) = \min(\tau_1, \dots, \tau_N),\ T_{\max}(h) = \max(\tau_1, \dots, \tau_N)$$

that will be referred to as minimal LD-CUSUM (Min-LD-CUSUM) and maximal LD-CUSUM (Max-LD-CUSUM) tests, respectively.

It can be shown that

$$\mathrm{ARL}(T_{\max}) \geqslant e^h \quad \text{and} \quad \mathrm{ARL}(T_{\min}) \geqslant N^{-1}e^h.$$

and that, as $h\to\infty$,

$$\mathrm{SADD}(T_{\min}) \sim \frac{h}{\max_i \mathrm{I}_i}, \quad \mathrm{SADD}(T_{\max}) \sim \frac{h}{\min_i \mathrm{I}_i},$$

Therefore, taking the thresholds $h = \log\gamma$ in the first case and $h = \log(N\gamma)$ is the second case, we obtain the tradeoff curves that relate the SADD and the ARL, as $\gamma\to\infty$:

$$\mathrm{SADD}(T_{\min}) \sim \frac{\log\gamma}{\max_i \mathrm{I}_i}, \quad \mathrm{SADD}(T_{\max}) \sim \frac{\log\gamma}{\min_i \mathrm{I}_i}.$$

It follows that in the symmetric case where $\mathrm{I}_i = \mathrm{I}$ the asymptotic relative efficiency of these detection tests compared to the optimal centralized test is

$$\mathrm{ARE}(T_{\min}; \tau_c) = \mathrm{ARE}(T_{\min}; \tau_c) = N.$$

4.4.4. Monte Carlo Experiments. In this section, we present the results of MC experiments for the Poisson example where observations in the i-th sensor $X_i(n)$, $n \geqslant 1$ follow the common Poisson distribution $\mathcal{P}(\mu_i)$ in the pre-change mode and the common Poisson distribution $\mathcal{P}(\theta_i)$ after the change occurs, i.e., for $m = 0, 1, 2, \dots$ and $\lambda = k$,

$$\mathbf{P}_k(X_i(n) = m) = \begin{cases} \frac{(\mu_i)^m}{m!} e^{-\mu_i} & \text{for } k > n, \\ \frac{(\theta_i)^m}{m!} e^{-\theta_i} & \text{for } k \leqslant n, \end{cases}$$

where without loss of generality we assume that $\theta_i > \mu_i$.

Write $Q_i = \theta_i / \mu_i$. It is easily seen that the LLR statistic in the i-th senor has the form

$$Z_n(i) = X_i(n) \log(Q_i) - \mu_i(Q_i - 1), \tag{14}$$

and the K-L information numbers

$$\mathrm{I}_i = \theta_i \log Q_i - \mu_i(Q_i - 1), \quad i = 1, \dots, N. \tag{15}$$

It follows from (5), (15) and the above discussion that the centralized CUSUM and AO-LD-CUSUM tests with the thresholds $h = \log \gamma$ are first-order globally asymptotically optimal and

$$\inf_{\tau \in \mathbf{\Delta}(\gamma)} \mathrm{SADD}(\tau) \sim \mathrm{SADD}(\tau_c) \sim \mathrm{SADD}(T_{ld}) \sim \frac{\log \gamma}{\sum_{i=1}^N [\theta_i \log Q_i - \mu_i(Q_i - 1)]}. \tag{16}$$

This means that the ARE of these detection tests with respect to the globally optimal test is equal to 1.

In order to evaluate the ARE of an optimal test ν (e.g., the centralized CUSUM test τ_c) with respect to the BQ-CUSUM test (8) we use (10), which yields

$$\mathrm{ARE}(\nu; \tau_b) = \frac{\sum_{i=1}^N \max_{t_i} [\beta_i(t_i) a_i(t) + a_{0,i}(t_i)]}{\sum_{i=1}^N [\theta_i \log Q_i - \mu_i(Q_i - 1)]}, \tag{17}$$

where the probabilities $\beta_{0,i}(t)$ and $\beta_i(t)$ are given by:

$$\beta_{0,i}(t_i) = \sum_{k=\lceil t_i \rceil}^{\infty} \frac{\mu_i^k \, e^{-\mu_i}}{k!}, \ \beta_i(t_i) = \sum_{k=\lceil t_i \rceil}^{\infty} \frac{\theta_i^k \, e^{-\theta_i}}{k!}.$$

The optimal values of $t_i^0 = \arg\max \mathrm{I}_i^b(t_i)$ that maximize the K-L numbers are easily found based on these formulas. Consider a symmetric case where $\mu_i = 10$ and $\theta_i = 12$ for all $i = 1, \dots, N$. Then $\mathrm{I}_i = \mathrm{I} = 0.1879$, the optimum threshold is $t_i^0 = 12$, and the corresponding maximum K-L distance for the binary sequence $\mathrm{I}_i^b(t_i^0) = \mathrm{I}^b = 0.119$. Therefore, the loss in efficiency of the BQ-test compared to the globally asymptotically optimal detection procedure is $\mathrm{ARE}(\nu; \tau_b) = 0.119/0.1879 = 0.63$, i.e., for the large ARL we expect about 37% increase in the average detection delay compared to the centralized CUSUM (C-CUSUM). The following MC simulations show that for the practically interesting values of the ARL (up to $13,360$) the gain of

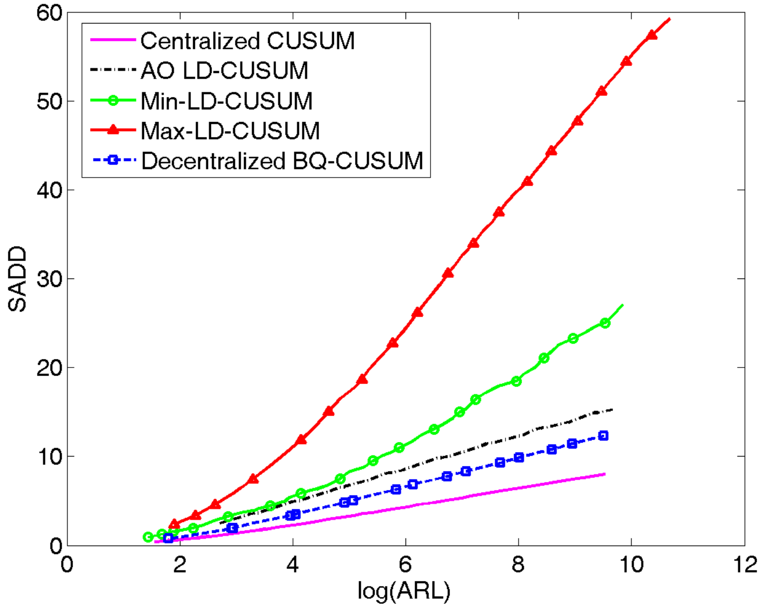


Figure 8: Operating characteristics of detection procedures

the optimal C-CUSUM test is even smaller, while the AO-LD-CUSUM test performs worse than the BQ-CUSUM test due to the reasons discussed in Section 4.4.3.

MC simulations have been performed for the above symmetric situation (i.e., $\mu_i = \mu = 10$ and $\theta_i = \theta = 12$) with $N = 5$ sensors. We used 10^5 MC replications in the experiment. The operating characteristics of the five detection tests (SADD vs $\log(\mathrm{ARL})$) are shown in Figure 8. It is seen that the BQ-CUSUM test substantially outperforms the AO-LD-CUSUM test for all false alarm rate range used in simulations. This result confirms our conjecture. It is also seen that both Min-LD-CUSUM and Max-LD-CUSUM perform worse than both BQ-CUSUM and AO-LD-CUSUM tests.

4.4.5. Conclusions and Future Work. The presented results allow us to compare performance of four proposed decentralized change detection procedures, as well as to determine loss in efficiency compared to the globally optimal centralized scheme. The first detection test, called the BQ-CUSUM test, uses binary quantizers at the sensors followed by the CUSUM detection procedure at the fusion sensor. The second detection test, called the AO-LD-CUSUM test, performs local detection at the sensors using CUSUM tests, and at each sampling point transmits these local decisions to the fusion center for combining and making the final decision. Both decentralized detection procedures transmit only binary sequences of 1's and 0's to the fusion center. Therefore, both detection tests use maximal possible level of data compression and require minimum bandwidth for communication. The third and fourth decentralized detection procedures, called the minimal and maximal LD-CUSUM tests respectively, are based on independent voting of sensors. In the former one the decision is made at the first time when the first CUSUM test detects the change; while in the latter one when all the sensors detect the change (but independently, not like in the AO-LD-CUSUM).

Due to losses of information, the BQ-CUSUM test is inferior to the globally optimal centralized CUSUM test. On the other hand, the AO-LD-CUSUM test is first-order asymptotically globally optimal for low false alarm rate. However, convergence to the optimum is expected to be slow, since the second term in the decomposition for the average detection delay goes to infinity as the square root of the threshold. We therefore conjectured that despite the fact that the AO-LD-CUSUM test is first-order asymptotically optimal it may perform worse than the non-optimal BQ-CUSUM test in realistic environment. The results of MC simulations for the Poisson model confirm this latter hypothesis. For the model considered the BQ-CUSUM outperforms the LD-CUSUM for all range of tested ARLs, from 33 to 13,360. The increase in the SADD is 30% for high false alarm rate and it slowly reduces to 18% for low false alarm rate. While potentially the ARE of the AO-LD-CUSUM test compared to the BQ-CUSUM test is 37%, this performance never kicks in for realistic moderately low false alarm rate.

The "voting" Min-LD-CUSUM and Max-LD-CUSUM tests are neither asymptotically optimal nor very efficient. Both tests are inferior to AO-LD-CUSUM and BQ-CUSUM tests. The Min-LD-CUSUM test is inferior to the Max-LD-CUSUM test in the symmetric case, and it is expected to perform even better in asymmetric scenarios.

The additional advantage of the BQ-CUSUM test compared to all other decentralized LD-CUSUM tests is that it does not require any processing power at the sensors.

While the considered Poisson model is motivated by network security applications such as rapid detection of computer intrusions, in reality it never holds and therefore efficient nonparametric detection procedures are needed. Suitable procedures will be developed during the Year 2 effort. Their comprehensive study (theoretical, MC simulations, and implementation for real data sets) for multi-sensor distributed systems is an important task of the future work. See, however, Section 4.7 for some preliminary results.

4.5. Energy-Efficient Tracking in Sensor Networks

4.5.1. Problem Description. The essential features of the tracking problem described in Figure 9 are contained in a one-dimensional simplification where the sensors are placed on a line and the object undergoes a random walk on the line. We hence consider this simplification in the sequel to facilitate presentation, with the understanding that the techniques that we develop can be generalized to the two-dimensional tracking problem.

Consider a one-dimensional sensor network with sensors placed unit distance apart from $-m$ to $+m$. An object that has to be tracked by this sensor network is assumed to undergo a random walk along the line. Let b_k denote the location of the object at time k. Then

$$b_{k+1} = b_k + w_k \tag{18}$$

where $\{w_k\}$ are i.i.d. integer-valued random variables with known distribution. We assume that $w_k \in [-n, n]$ with n typically being much smaller than m. For example, $\{w_k\}$ could be i.i.d. Bernoulli random variables that take the value $+1$ or -1 with equal probability. The tracking problem stops when the object leaves the network, i.e., when $b_k \notin \{-m, \ldots, 0, \ldots, m\}$.

A central unit, which controls this sensor network, is assumed to maintain the information required to compute the sleep times of the sensors in the system and to assign the sleep times for

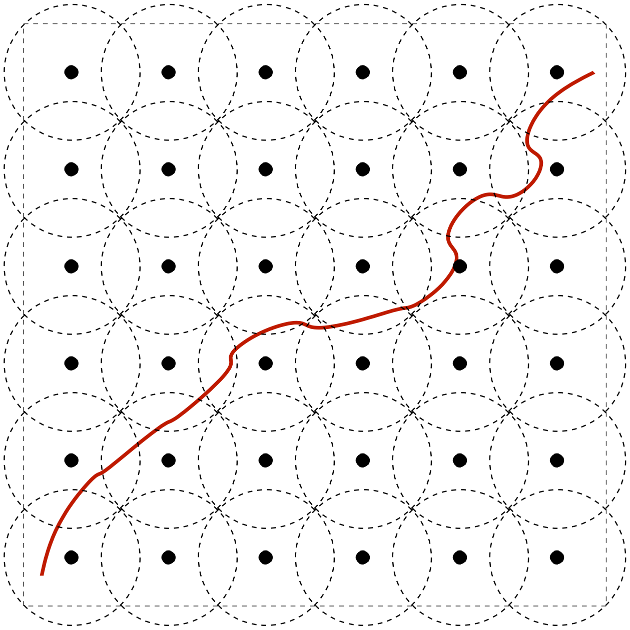


Figure 9: Object tracking in a field of sensors

the sensors that come awake. A sensor is either awake or asleep at each time instant. Each sensor that wakes up remains awake for one time unit during which the following actions are taken: (i) if the object is within its range, the sensor detects the object and sends this information to the central unit, and (ii) the sensor receives a new sleep time (which may equal zero) from the central controller. The input from the central unit is used to set a sleep timer at the sensor, which gets decremented by one every time unit.

Let $r_{k,\ell}$ denote the residual sleep time at time k for the sensor located at position ℓ, i.e., $r_{k,\ell}$ is the value of the sleep timer at sensor ℓ at time k. Also let $u_{k,\ell}$ denote the control input (sleep time) given to sensor ℓ from the central unit at time k. We can write the update of $r_{k,\ell}$ as

$$r_{k+1,\ell} = (r_{k,\ell} - 1)\mathbb{1}_{\{r_{k,\ell}>0\}} + u_{k,\ell}\mathbb{1}_{\{r_{k,\ell}=0\}} \tag{19}$$

where $\mathbb{1}$ is the indicator function. We use the vector notation $\boldsymbol{r}_k = (r_{k,-m}, \ldots, r_{k,m})$ and $\boldsymbol{u}_k = (u_{k,-m}, \ldots, u_{k,m})$.

Based on (18) and (19), we see that we have discrete-time dynamical model that describes our tracking problem, with exogenous input w_k and control input $\boldsymbol{u}_k$. The *state* of the system at time k is described by $x_k = (b_k, \boldsymbol{r}_k)$ and it has the following evolution in time:

$$x_{k+1} = \begin{cases} f(x_k, \boldsymbol{u}_k, w_k) & \text{if } x_k \neq \mathcal{T} \\ \mathcal{T} & \text{if } x_k = \mathcal{T} \text{ or if } b_k \notin \{-m, \ldots, m\} \end{cases} \tag{20}$$

where $\mathcal{T}$ denotes a terminal state that the system reaches when the objects exits the sensor network, and f is described by (18) and (19). Once in the terminal state, the system remains there indefinitely. With some possible abuse of notation, we denote the components of the terminal state corresponding to both b_k and $\boldsymbol{r}_k$ by $\mathcal{T}$.

Unfortunately, not all of x_k is known to the central unit at time k since b_k is known only if the sensor at location b_k is awake at time k. Thus we have dynamical system with incomplete (or partially observed) state information. If we denote the observation available to the central unit at time k by z_k, then $z_k = (s_k, \boldsymbol{r}_k)$, with

$$s_k = \begin{cases} b_k & \text{if } b_k \neq \mathcal{T} \text{ and } r_{k,b_k} = 0 \\ \mathcal{E} & \text{if } b_k \neq \mathcal{T} \text{ and } r_{k,b_k} > 0 \\ \mathcal{T} & \text{if } b_k = \mathcal{T} \end{cases}$$

where $\mathcal{E}$ denotes an unknown or "erasure" value.

The total information available to the control unit at time k is given by

$$I_k = (z_0, \ldots, z_k, \boldsymbol{u}_0, \ldots, \boldsymbol{u}_{k-1}). \tag{21}$$

with $I_0 = z_0$ denoting the initial (known) state of the system. The control input for sensor ℓ at time k is allowed to be a function of I_k, i.e.,

$$u_{k,\ell} = \mu_{k,\ell}(I_k).$$

We assume that an energy cost of unity is contributed by each sensor that is awake, and a tracking cost of c is incurred for each time unit that the object is not tracked. The total cost at time k is then given by

$$g(x_k) = \mathbb{1}_{\{x_k \neq \mathcal{T}\}} \left[c\, \mathbb{1}_{\{r_{k,b_k} > 0\}} + \sum_{\ell=-m}^{m} \mathbb{1}_{\{r_{k,b_k} = 0\}} \right]. \tag{22}$$

Thus c is the parameter used to tradeoff energy consumption and tracking errors, and the total cost values for different values of c produce the tradeoff curve for a given sleeping policy.

The total cost (over a possibly infinite horizon trajectory) for the system is given by

$$J_0(I_0, \mu_0, \mu_1, \ldots) = \mathsf{E} \left[\sum_{k=1}^{\infty} g(x_k) \middle| I_0 \right]$$

Since g is bounded by $(2m + 1 + c)$, the cost function J_0 is guaranteed to be bounded as long as the expected time for the object to exit the system is finite. The latter condition holds for any nontrivial random walk. Hence the following optimization problem is well defined.

$$J_0^*(I_0) = \min_{\mu_0, \mu_1, \ldots} J_0(I_0, \mu_0, \mu_1, \ldots) \tag{23}$$

The solution to this optimization problem for each value of c yields an optimal sleeping policy. The optimization problem falls under the framework of partially observable Markov decision process (POMDP), and the optimal solution may be obtained via dynamic programming (DP).

4.5.2. Optimal Solution via Dynamic Programming. We begin with identifying sufficient statistics for the tracking problem.

Sufficient statistic for DP. The information for decision-making at time k given in (21) is unbounded in memory. It is easy to show via standard arguments (see, e.g. [3]) that a sufficient statistic for optimization, that is bounded in memory, is given by the probability distribution of the state x_k, given I_k. Since $\boldsymbol{r}_k$ is part of x_k, the sufficient statistic can be written as $v_k = (\boldsymbol{r}_k, \boldsymbol{p}_k)$, where $\boldsymbol{p}_k$ is a row vector that denotes the probability distribution of the location of the object, b_k, given I_k. The components of $\boldsymbol{p}_k$ are given by:

$$p_{k,\ell} = \mathsf{P}(\{b_k = \ell\}|I_k), \quad \ell = -m, \ldots, m, \tag{24}$$

and $p_{k,m+1} = \mathsf{P}(\{b_k = \mathcal{T}\}|I_k)$.

The sufficient statistic (or belief state as it is referred to in the POMDP literature [1]) can be updated recursively based on the new observation. It is easiest to see this in two steps. First we update $\boldsymbol{p}_k$ without using the new observation z_{k+1}, i.e., using only I_k to form vector $\boldsymbol{q}_{k+1}$ with components

$$q_{k+1,\ell} = \mathsf{P}(\{b_{k+1} = \ell\}|I_k) \tag{25}$$

and $q_{k+1,m+1} = \mathsf{P}(\{b_{k+1} = \mathcal{T}\}|I_k)$. The vector $\boldsymbol{q}_{k+1}$ is obtained from $\boldsymbol{p}_k$ via a Markov evolution with transition matrix $\mathbb{P}$ defined by statistics of the jump variables $\{w_k\}$:

$$\boldsymbol{q}_{k+1} = \boldsymbol{p}_k \, \mathbb{P} \tag{26}$$

The last row of $\mathbb{P}$ corresponds to the absorbing terminal state.

We now "clean up" $\boldsymbol{q}_{k+1}$ using the new observation z_{k+1} as follows. If the object is observed at sensor ℓ, we replace $\boldsymbol{q}_{k+1}$ with a unit point mass at ℓ. If the object is not observed by any of the sensors that are awake, we zero out the those components of $\boldsymbol{q}_{k+1}$ and normalize the remaining ones. Thus

$$p_{k+1,\ell} = \mathbb{1}_{\{s_{k+1}=\ell\}} + \mathbb{1}_{\{s_{k+1}=\mathcal{E}\}} \, \mathbb{1}_{\{r_{k+1,\ell}\neq 0\}} \frac{q_{k+1,\ell}}{\sum_i \mathbb{1}_{\{r_{k+1,i}\neq 0\}} q_{k+1,i}}. \tag{27}$$

Tractability of optimal solution. We can easily write down the finite-horizon DP equations in terms of the sufficient statistic $v_k = (\boldsymbol{r}_k, \boldsymbol{p}_k)$. Furthermore, it is easily established that the finite-horizon cost-to-go functions converge as the horizon goes to infinity and that the corresponding limits are independent of k due to the stationary nature of the problem. Thus the optimal cost in (23) is given by the infinite-horizon cost-to-go function, and the corresponding optimal control functions μ_k are the same for all k. The optimal cost and the optimal sleeping policy can hence be found by solving a Bellman equation [3], via known techniques such as successive approximation. However, the optimal solution is intractable even for small sensor networks. This is because the state space grows exponentially with the number of sensors. For example, even with seven sensors with maximum sleep time of only 10 and probability mass function quantized to multiples of 0.1, there are about 10^9 possible states v_k.

4.5.3. Practical approximations. We now address the problem of finding practical tractable solutions, yet efficient in terms of statistical performance.

Q_{MDP} solution. Because the optimal solution is intractable, we wish to formulate an alternative problem that is tractable yet retains most of the essential features of the optimal solution. A popular approach to finding good suboptimal solutions for POMDP's is to assume that at times after the current time, we will have perfect state information. The solution so obtained is known in the literature as the $\mathrm{Q_{MDP}}$ solution [1].

We assume that beyond the current time, each sensor somehow knows the exact position of the object each time it wakes up. Thus, whenever a sensor wakes up, the set of possible distributions it sees is the set of point mass distributions. Under this assumption it is clear that from the perspective of a sensor ℓ, the actions of the other sensors do not affect the state evolution. We also know that a sensor ℓ can only affect the cost that accrues either when sensor ℓ comes awake or when a tracking error occurs at sensor ℓ. Thus, the optimization problem under this assumption fully separates into $2m+1$ problems — one for each sensor.

Let us solve the optimization problem at sensor ℓ using an infinite-horizon dynamic program. Since the residual sleep times of the other sensors are irrelevant to optimal decision making in the $\mathrm{Q_{MDP}}$ setting, the sufficient statistic for decision making at time k is simply $\boldsymbol{p}_k$. The Bellman equation for this problem is easily shown to be

$$J^{(\ell)}(\boldsymbol{p}) = \min_{\mu^{(\ell)}} \Big(\sum_{i=1}^{u} c\left[\boldsymbol{p}\mathbb{P}^i\right]_\ell + \sum_{j\neq\mathcal{T}} \left[\boldsymbol{p}\mathbb{P}^{u+1}\right]_j + \sum_k \left[\boldsymbol{p}\mathbb{P}^{u+1}\right]_k J^{(\ell)}(\boldsymbol{e}_k) \quad \Big)\Big|_{u=\mu^{(\ell)}(\boldsymbol{p})} \tag{28}$$

where $\boldsymbol{e}_b$ denotes a row vector with a one in position b and zeros everywhere else, and where $J^{(\ell)}$ is the infinite-horizon cost-to-go function for sensor ℓ. The $\mathrm{Q_{MDP}}$ policy for sensor ℓ, $\mu_Q^{(\ell)}$, is given from the minimization on the RHS of (28).

If we can solve (28) for $\boldsymbol{p} = \boldsymbol{e}_b$ for $b \in \{-m, \ldots, m, \mathcal{T}\}$, we have sufficient information to define the solution for any other distribution $\boldsymbol{p}$. Thus we have $2m+2$ equations in $2m+2$ unknowns. However, this set of equations does not have a unique solution since we can add an arbitrary constant to a solution $J^{(\ell)}$ and still satisfy the equations. We therefore add the additional constraint that $J^{(\ell)}(\boldsymbol{e}_\mathcal{T}) = 0$ (which is clearly the desired solution). This reduces the problem to one of $2m+1$ equations in $2m+1$ unknowns with a unique solution. An effective method for finding the solution is to use policy iteration [1].

A lower bound on optimal performance. Since the $\mathrm{Q_{MDP}}$ solution assumes more information than is actually available, the cost obtained in its derivation is a lower bound on the cost of any scheme. In particular, if we apply the $\mathrm{Q_{MDP}}$ policy to the actual system (without perfect state information), we will achieve a higher cost. Intuitively, the lower bound should be tightest when the number of tracking errors is small so that the assumption that the position of the object is known is most realistic.

Point mass approximations. The $\mathrm{Q_{MDP}}$ policy, $\mu_Q = \{\mu_Q^{(\ell)} : \forall \ell\}$ is considerably easier to compute than the optimal policy and can be computed on-line after some initial off-line computation has been completed. However, such on-line computation requires sufficient processing power and could introduce delays. It would be convenient if $\mu_Q^{(\ell)}$ could be pre-computed and stored either at the central controller or at sensor ℓ itself. The latter option is particularly attractive since it allows for decentralized implementation. But the set of possible distributions $\boldsymbol{p}$ is potentially quite large — even if quantization is performed — and could make the storage requirements prohibitive.

To make the storage requirements feasible, we consider approximations of the Q_{MDP} algorithm where $\boldsymbol{p}$ is replaced by a unit point mass distribution. There are two options for the placement of the unit point mass: (i) the centroid of $\boldsymbol{p}$, and (ii) the nearest point to the sensor on the support of $\boldsymbol{p}$. The latter option allows for the implementation of the Q_{MDP} policy without detailed information about the statistics of the random walk – only the support of the jump variables w_k is required!

4.5.4. Numerical Results. Simulations of the various polices were performed for 1-D sensor networks. In these simulations, the object was initially placed at the center of the network and the location of the object was made known to each sensor. By averaging over many simulation runs, it was possible to compute the average number of tracking errors and the average number of sensors awake per unit time. These values could then be plotted for different values of c to generate a tradeoff curve for these two quantities.

Figures 10 and 11 show results for two different networks. The results of Figure 10 are for a network with 41 sensors ($m = 20$) where the object moved according to a symmetric random walk. In other words, the $\{w_k\}$ were i.i.d. random variables taking on value $+1$ or -1 with equal probability. The results of Figure 11 are for a network with 61 sensors ($m = 30$) where the $\{w_k\}$ were i.i.d. random variables uniformly distributed over $\{-3, -2, \dots, 2, 3\}$.

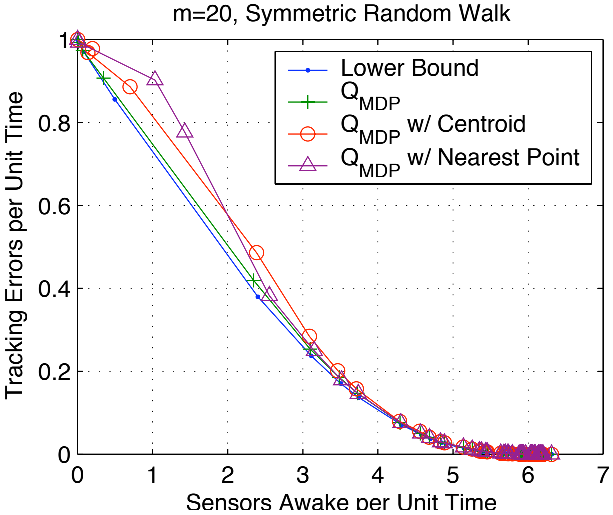


Figure 10: Comparison of lower bound and Q_{MDP} solutions for $m = 20$

Four curves are plotted in each figure. The first curve is the tradeoff curve that results from the lower bound described in the previous section. Although this curve is unachievable, it is useful as a baseline since if a sleeping policy approaches the performance of this reference curve, that sleeping policy must also be approaching optimal performance. The remaining three curves are simulation results for the Q_{MDP} solution and for the Q_{MDP} solution using the centroid and nearest point approximations described in the previous section.

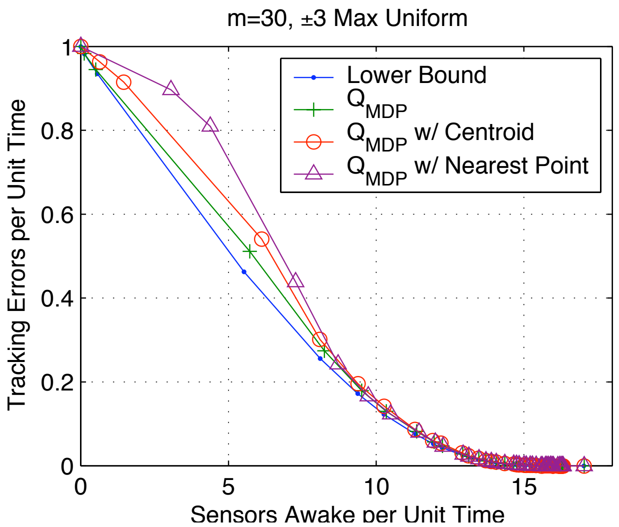


Figure 11: Comparison of lower bound and $\mathrm{Q_{MDP}}$ solutions for $m = 30$

From these simulation results, we see that the $\mathrm{Q_{MDP}}$ solution is very close to the curve for our lower bound and is thus nearly optimal. This is especially true in the regime of interest where the tracking error is small. The use of point mass approximations does result in some loss of performance, but again this loss is small for small tracking error.

Note that we could also consider a more primitive policy (which does not use location information) where each sensor would be awake with some probability π at each time instant. As π were varied, we would achieve a tradeoff curve that is a straight line between the points $(0, 1)$ and $(2m + 1, 0)$ in the coordinate system used in the above plots. When compared with this tradeoff curve, the schemes we have proposed result in significant improvement.

We have obtained similar results for a variety of other cases for the object trajectory, including one-dimensional walks with more complicated statistics for w_k, and two-dimensional random walks, which we could not present here due to space limitations. We have also designed a simpler suboptimal solution called the FCR solution, which suffers from some performance loss relative to the $\mathrm{Q_{MDP}}$ solution. The details are given in [11]. Extensions to more realistic object movement and sensing models are described in [12].

4.6. Information Integration and Fusion in Distributed Heterogeneous Multisource Multisensor Systems

4.6.1. Multi-Vehicle Motion and Sensing. In [7], we study cooperative control algorithms using pairwise interactions, for the purpose of controlling flocks of unmanned vehicles. An important issue is the role the potential plays in the stability and possible collapse of the group as agent number increases. We model a set of interacting Dubins vehicles with fixed turning angle and speed. We perform simulations for a large number of agents and we show experimental realizations of the model on a testbed with a small number of vehicles. In both cases, critical thresholds exist between coherent, stable, and scalable flocking and dispersed or collapsing motion of the group.

The paper [19] describes the second generation of an economical cooperative control testbed described in [15]. The original car-based vehicle is improved with on-board range sensing, limited on board computing, and wireless communication, while maintaining economic feasibility and scale. A second, tank-based platform, uses a flexible caterpillar-belt drive and the same modular sensing and communication components. We demonstrate practical use of the testbed for algorithm validation by implementing a recently proposed cooperative steering law involving obstacle avoidance.

Autonomous robotic systems (observers) equipped with range sensors must be able to discover their surroundings, in an initially unknown environment, for navigational purposes. In [18], we present an implementation of a recent environment mapping algorithm [17] based on Essentially Non-oscillatory (ENO) interpolation [14]. The tank-based platform is used to validate our algorithm due to the ability of the tanks to turn in place. The tanks are equipped with a flexible caterpillar drive, range sensor, limited onboard computing, and wireless communication. This project was jointly sponsored by Los Alamos National Lab and the Research in Industrial Projects for Students at the Institute for Pure and Applied Mathematics.

4.6.2. Change-Point Detection Methods for Obstacle Avoidance. We implemented a basic change-point detection method on the platform described above, using the car-based chassis rather than the tank. The particular task was obstacle avoidance in real time using a noisy IR range sensor mounted on the front of the vehicle. As the vehicle approaches the obstacle, sensor readings adjust from background noise to a level indicating the presence of the object. Figure 12 shows an example of raw sensor readings. To filter the signal, we use a particular version [27] of a standard cumulative sum algorithm [2]. Let X_n denote the raw sensor signal at time level n and μ denote the mean of the background noise when no obstacle is present. Define $Z_n = X_n - \mu - c$ where c is a fraction of the expected change in sensor reading due to the obstacle. Next define recursively $W_n = \max(0, Z_n + W_{n-1})$, $n = 1, 2, \dots$ $(W_0 = 0)$. The calculated value W_n should remain around zero until the change of state occurs, at which point it ramps up. An example is shown in Figure 12. Once W_n passes a designated threshold (large enough to avoid false alarms with a high probability) the object is detected. Using the car chassis at 1/5 of the full throttle, we test the cumulative sum algorithm for different values of c ranging from 150 to 400. The results are well-reproduced in multiple trials. These values lie closely on a linear fit, therefore we use the $c = 200$ state in practice for the most advanced warning.

The change point detection algorithm allowed a team of vehicles to avoid an obstacle while reaching a target on the other side of the obstacle. The result is shown in Figure 13.

4.6.3. Spatio-Temporal Image Segmentation and Video Tracking. We consider the complex task of video tracking under occlusions and in complex backgrounds. We assume that the complete object boundary is known through prior-shape template information and that the object undergoes only affine motion across frames. Our approach is to track the object by segmenting it in each frame, and also simultaneously registering it with the given template. The method is based a level set idea [6, 20] but also uses a logic based (OR/AND) segmentation framework to segment the object under occlusions and complex backgrounds. The novelty of this method is our automatic choice of the OR/AND logic model based on prior shape information. The developed tracking algorithm under partial occlusions utilizes Logic Models with the addition of prior shape information. We represent object motion as a registration between frames. We can track successfully as

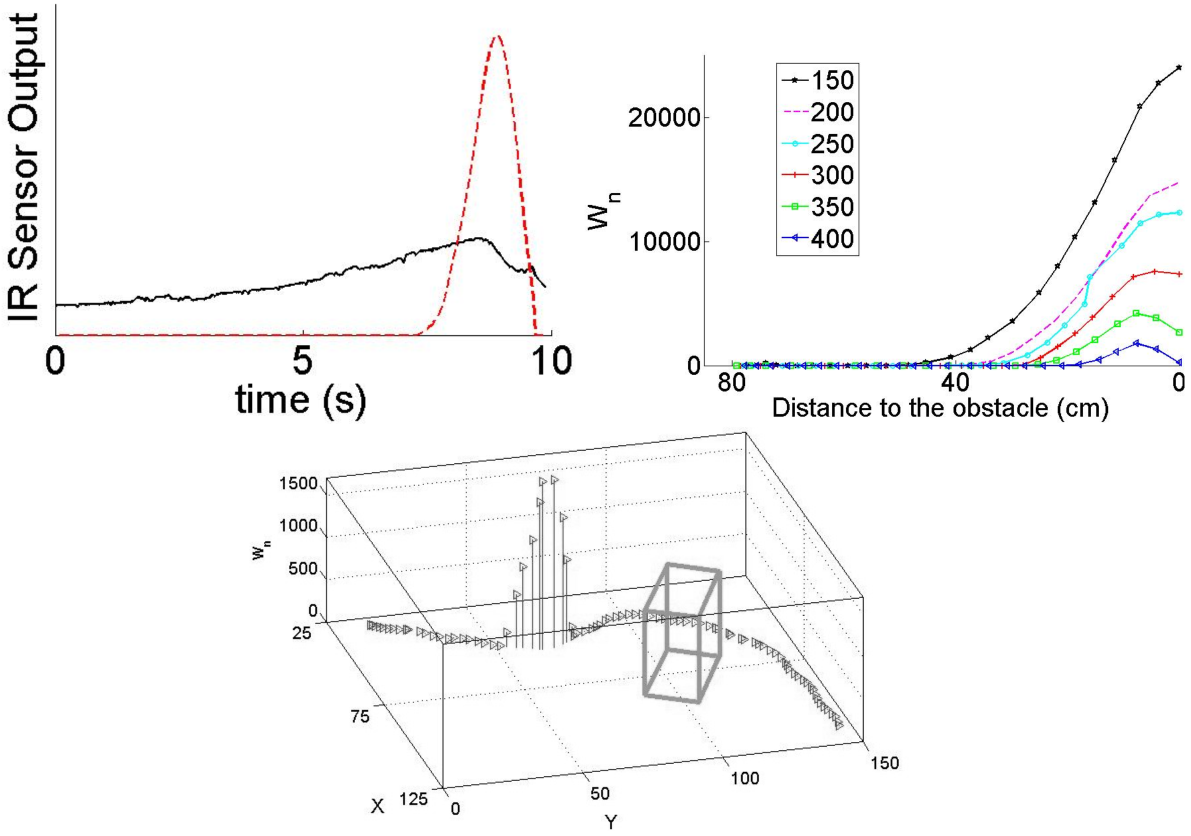


Figure 12: Cumulative sum algorithm applied to sensor data from a single car approaching an obstacle. (top left) raw data and cumulative sum, (top right) cumulative sums for different choices of c, (bottom) sample car path avoiding obstacle with cumulative sum sensor output.

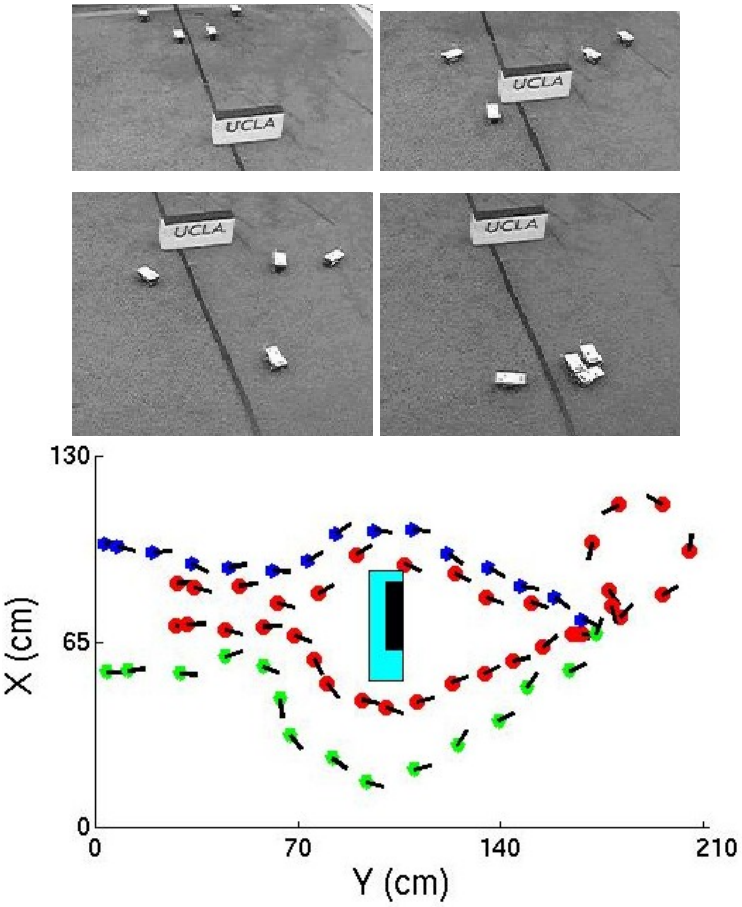
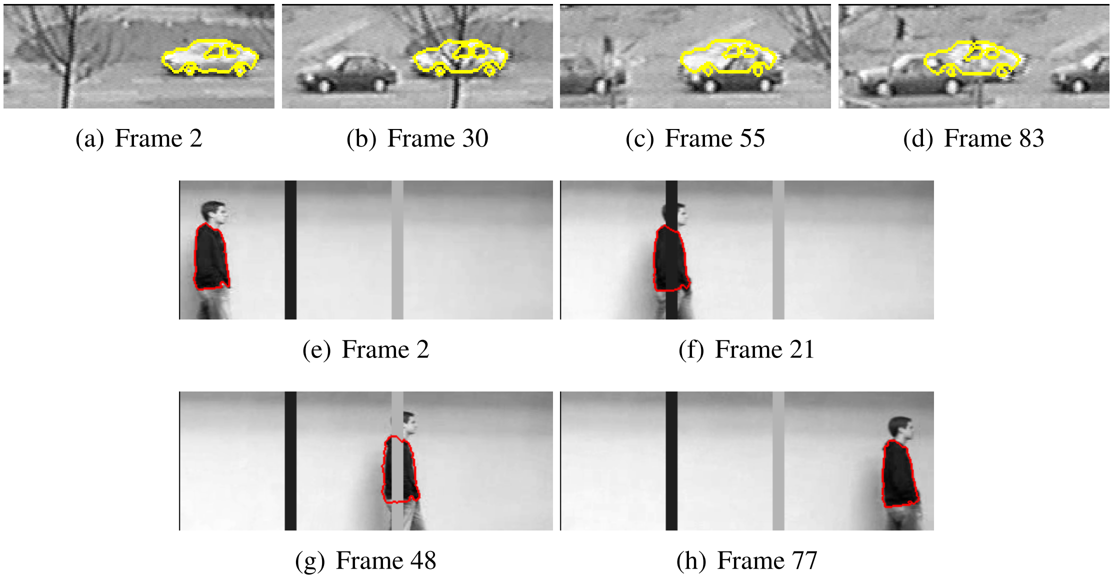


Figure 13: Target seeking with barrier avoidance. Top four panels show snapshots, at different times, of a single demonstration of the maneuver. The time progresses from top left to bottom right. The bottom figures shows trajectories of the cars compared to both the actual barrier (dark) and the larger virtual barrier (light) as computed from the range sensors of the observers.

long as the object of interest maintains nearly constant shape and intensity throughout the sequence, and does not become totally occluded. In addition the algorithm provides object contouring. See [4] for more information.

The results of tracking are shown in Figure 14.



(a) Frame 2 (b) Frame 30 (c) Frame 55 (d) Frame 83

(e) Frame 2 (f) Frame 21

(g) Frame 48 (h) Frame 77

Figure 14: Occlusion tracking from two video sequences

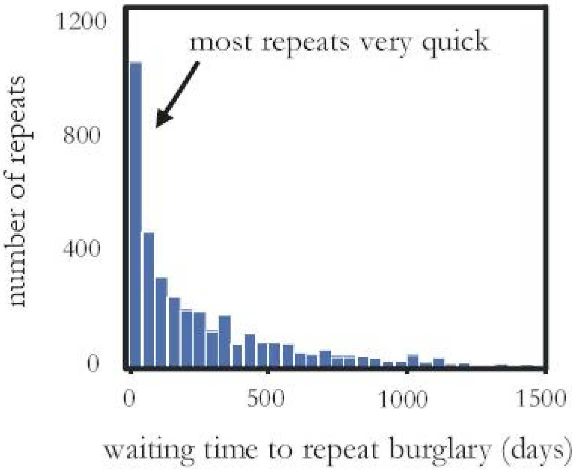


Figure 15: Residential burglaries in the city of Long Beach, 2000-2005. Number of repeat break ins at the same location, as a function of time from original break in

.

4.6.4. Models for Crime Patterns. We worked on models for residential burglaries. The goal was to show how crime opportunities and motivated offenders can lead to unevenly distributed events. Basic foraging strategies are what bring motivated offenders together with criminal opportunities. Of particular interest in this study was the fact that repeat burglaries at the same location follow a rapid temporal drop off (see Figure 15).

We built a mathematical model for these events and have shown that this leads to spatial crime hotspots in a neighborhood.

4.7. Applications of CPD and Spectral Signal Processing Methods to Intrusion Detection in Distributed Computer Networks

One of the important applications of distributed change-point detection methods developed in Section 4.4 is intrusion detection in distributed high-speed computer networks. A significant number of serious cyberattacks on a variety of governmental agencies, universities, and corporations have recently been identified [16]. These attacks, including a variety of buffer overflows, worm-based, denial-of-service (DoS) and man-in-the-middle (MiM) attacks, are designed to gain access to additional hosts, steal sensitive data, and disrupt network services. As a result, *rapid detection* of a wide spectrum of network intrusions and *robust separation* of legitimate and malicious traffic are vital for the continuation of normal operation of networks. See Kent [16] and Tartakovsky et all [27]-[29] for a more detailed discussion.

Typically network intrusions occur at unknown points in time and lead to changes in the statistical properties of certain observables. For example, distributed DoS (DDoS) attacks lead to changes in the mean value of the number of packets of a particular type (TCP, ICMP, or UDP) and size, while address resolution protocol (ARP) MiM attacks lead to changes in the average number of ARP requests [27]-[29]. It is therefore intuitively appealing to formulate the problem of detecting attacks as a quickest change-point detection problem: to detect changes in statistical models as rapidly as possible (i.e., with minimal average delays) while maintaining the false alarm rate at a given low level.

4.7.1. Nonparametric Distributed Change Detection Algorithms for Detecting Intrusions. It follows from the results of Section 4.4 that in the case of complete information about the pre-change and the post-change models, (asymptotically) optimal detection procedures in multisensor detection systems can be constructed based on the LLR-based CUSUM tests. However, in intrusion detection applications, these models are unknown. For this reason, in [27]-[29], a nonparametric approach was proposed and thoroughly tested for a single-sensor scenario. This approach can be extended easily to the multisensor centralized and decentralized scenarios.

More specifically, when the pre-change and post-change densities are unknown, the LLRs $Z_i(n)$ defined in (4) are also unknown and should be replaced by appropriate score functions $s_i(n)$ that have negative mean values $\mathbf{E}_\infty s_i(n) < 0$ before the change occurs and positive mean values $\mathbf{E}_k s_i(n) > 0$ after the change occurs.

While we do not specify any particular model in terms of probability distributions, some assumptions on the change should be made. Indeed, score functions can be chosen in many ways, and their selection depends crucially on the type of change that we intend to detect. For example, different score functions are used to detect changes in the mean and changes in the variance. In applications of interest, the detection problem can be usually reduced to detecting changes in mean values.

Let $\mu_i = \mathbf{E}_\infty X_i(j)$ and $\theta_i = \mathbf{E}_1 X_i(j)$ denote the pre-change and post-change mean values in the i-th sensor. Typically, the baseline mean values μ_i can be estimated quite accurately in advance while the values of θ_i are usually unknown and either should be estimated on-line or replaced by reasonable numbers, e.g., by the expected minimal values. In the rest of this subsection we suppose for concreteness that $\theta_i > \mu_i$.

For $i = 1, \dots, N$, introduce the following score functions $s_i(n) = X_i(n) - \mu_i - c_i$, where in the general case $c_i = c_i(n)$ may depend on past observations, which is desirable to guarantee an adaptive structure of the detection procedure. For example, one may take $c_i(n) = \varepsilon\hat{\theta}_{i,n}$, where ε is a tuning parameter belonging to the interval $(0, 1)$ and $\hat{\theta}_{i,n} = \hat{\theta}_{i,n}(\mathbf{X}_i^n)$ is an estimate of the unknown mean θ_i. Choosing the latter estimators as well as optimizing the parameter ε based on the training data are not straightforward tasks, as discussed in detail in Tartakovsky et al [27]. For this reason, it is convenient to set $c_i(n) = c_i$, where c_i are positive constants that do not depend on n.

Positiveness of c_i is essential to guarantee the negative value of $\mathbf{E}_\infty s_i(n) = -c_i$ under the no-change hypothesis. On the other hand, c_i does not have to be too large in order to guarantee the positive value of $\mathbf{E}_1 s_i(n) = \theta_i - \mu_i - c_i$ under the alternative hypothesis. A particular choice of c_i is discussed in [27].

If the above conditions hold, the score-based CUSUM statistic in the i-th sensor

$$W_i^s(n) = \max\{0, W_i^s(n-1) + s_i(n)\}$$

remains close to zero in normal conditions while when the change occurs it starts rapidly drifting upward (see Figure 16 for a typical behavior). The combined from all the sensors, centralized CUSUM statistic

$$W^s(n) = \max\left\{0, W^s(n-1) + \sum_{i=1}^{N} s_i(n)\right\}$$

has a similar behavior.

The time of alarm in the centralized detection scheme is defined as the first time n when the statistic $W^s(n)$ crosses a positive threshold.

A binary quantized version of the CUSUM test can be designed analogously to Section 4.4.2. See [27] for further details.

Finally, a nonparametric LD-CUSUM test has the form (11) where the LLR-based CUSUM statistic $W_i(n)$ is replaced with the score-based CUSUM statistic $W_i^s(n)$ and where

$$\pi_i = \frac{\theta_i - \mu_i - c_i}{\sum_{i=1}^{N}(\theta_i - \mu_i - c_i)}.$$

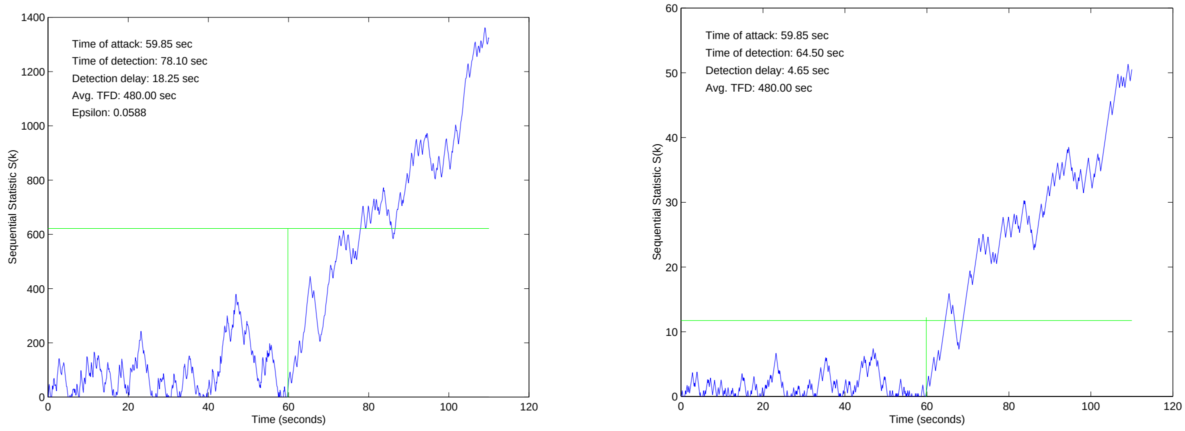
For the sake of simplicity, we assume here that the post-change mean values θ_i are known.

Note that the above nonparametric detection algorithms are no longer guaranteed to be optimal. Certain optimization is possible based on the training data [27].

4.7.2. Experimental Results: Rapid Detection of DDoS Attacks. We now present the results of experimental study of the distributed CPD algorithms proposed in the previous section for detecting intrusions in distributed computer networks. Specifically, we report the results of testing the NP-CUSUM and B-CUSUM procedures' abilities to detect a TCP SYN flooding attack based on real network traffic data collected and made available by the MIT Lincoln Laboratory.

The data sets in the study contain extensive training data that was collected during several sessions over a three-year period. We have used the 2000 LLDoS 1.0 intrusion scenario specific data sets and a 1999 training data set. The attack consisted of three hijacked servers sending many TCP connections to a victim host. The goal of the attack was to prevent other hosts from connecting to the server by overloading its resources and those of the network. The data set was manually split into two portions: one before an abrupt change in the traffic and one after. This was possible since the change was clearly visible. (We, however, cannot guarantee that the selected moment in time really was the moment of the attack.) In order to better illustrate the properties of the algorithms, we have made the change in traffic less obvious and the detection task more difficult. To this end, we have combined a training data set with normal traffic with the data set containing the assumed attack traffic and have modified the average number of packet arrivals per second by re-scaling the times between packets during the assumed attack. This technique can be considered as a simulated TCP flooding attack observed in real "background" traffic and gives us a flexibility in controlling attack intensities. Resampling of the traffic was used to estimate the performance of the algorithms.

We have considered two scenarios for detecting the change in the average number of observed packets. The length of the sampling period during which the packets were counted was 0.1 second in both scenarios. In Scenario 1, the mean value increased from 12.5 to 21.32 while the standard deviation of the observed packets remained almost the same before (17.96) and after the change (17.62). In Scenario 2, we made the detection task more difficult by simultaneously increasing the mean by a smaller amount from 12.5 to 16.04 and decreasing the standard deviation of the observed packet numbers from 17.96 to 13.54. Since the binary LR-CUSUM test is sensitive to changes not only in mean values but in variance as well, we expect that it performs best in the second scenario.

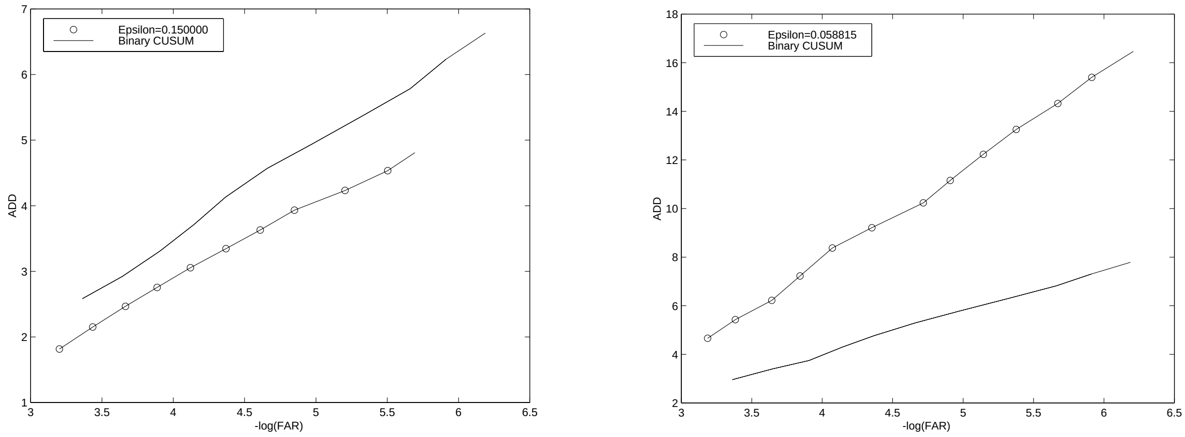


(a) Scenario 2: NP-CUSUM

(b) Scenario 2: Binary LR-CUSUM

Figure 16: NP-CUSUM and B-CUSUM detection statistics

The plots in Figure 16 illustrate the typical behavior of the detection statistics for the NP-CUSUM and B-CUSUM tests for Scenario 2. It is seen that the statistics fluctuate not very far from the zero reflection barrier for the legitimate traffic, but start rapidly drifting upward after the attack occurs. It is also seen that the B-CUSUM detection statistic has much less variability under the no-attack hypothesis as compared to that of the NP-CUSUM, which results in lower threshold



(a) Scenario 1

(b) Scenario 2

Figure 17: Operating characteristics of the NP-CUSUM and binary CUSUM tests

values for the same FAR. This is the second reason why we expect that the B-CUSUM test is more efficient in the second scenario than the NP-CUSUM test.

The NP-CUSUM test was first carefully optimized. The optimal value of ε was estimated as $\varepsilon_{\text{opt}} = 0.15$ in Scenario 1 and as $\varepsilon_{\text{opt}} = 0.058815$ in Scenario 2. The operating characteristics for both scenarios are shown in Figure 17. In Scenario 2, the graph illustrates the effect of decreasing the variance of the observed data after the change of the mean. In this situation, the thresholds of the NP-CUSUM tests must be higher than in the case where the variance remains the same. As we just mentioned, the binary test is sensitive not only to changes in mean values but also to variance changes; as a result, the binary test performs significantly better, as predicted above. In Scenario 1, the optimized NP-CUSUM test performs better, as can be expected from the preceding discussion.

4.7.3. Spectral Analysis Techniques for Signature-based IDS. The fundamental difficulty in detecting the attacks we described earlier is the inability of current, packet content-based IDS's to gather enough information about a packet stream to classify it as an attack. Specifically:

- Encrypted attacks do not allow access to application headers and payload to perform signature detection. Therefore, traditional content-based signatures cannot be extracted and distributed to other IDSs.
- Low-level attacks are hard to detect because they produce a very small signal near the source, which is buried in normal traffic. An IDS looking for such attacks would have to expend many resources examining numerous small flows, which would significantly hamper the operation of the IDS.
- Attacks through proxies present a challenge because malicious and legitimate traffic are mixed in a way that makes it hard for the IDS to distinguish them. For example, attack traffic coming from a Network Address Translation (NAT) proxy has the same source address as legitimate traffic, making it impossible to filter on source address.

Our approach is to use spectral analysis techniques applied on packet arrival time series to detect such attacks. Such techniques do not rely on packet content, and are thus impervious to encryption. Additionally, low-level and proxy attacks are now transformed into low-level signal detection in the aggregate, enabling us to use standard signal processing techniques for detection. Thus, instead of producing traditional content-based signatures, we can produce more robust spectral signatures, which can be distributed to other IDSs.

4.7.4. Hybrid Anomaly-Signature Intrusion Detection System. FAR of anomaly-based detectors with hard decisions may be improved by analyzing more detailed patterns in traffic statistics, i.e., signatures. Therefore, combining spectral signature approach and corresponding signal processing techniques with anomaly change detection based techniques seems to be beneficial. This approach is complementary to the anomaly-based and signature-based IDSs and allows for profiling, i.e., confirmation or rejection of detection decisions at the output of the anomaly detector using signature analysis. Combining these two methods into a hybrid IDS is not a trivial task.

To be more specific, the idea is to design a hybrid anomaly-signature IDS where the anomaly detector is followed by automatic signature analysis tools. Our definition of signatures is different from the conventional approach: flow-based signatures, e.g., spectral. See Section 4.7.3.

Approach summary:

- Hybrid algorithms with profiling capability that combine advanced statistical anomaly detection methods (such as change detection) with flow-based signature detection algorithms (such as spectral-based methods).
- This approach is complementary to the anomaly-based and signature-based IDSs and allows for profiling, i.e., confirmation or rejection of detection decisions at the output of the anomaly detector using signature analysis.
- Allows for lowering FAR, while keeping detection delays to a minimum.
- Allows for automatic signature generation and update.

5. POTENTIAL IMPACTS

The research will produce novel spatial-temporal nonlinear estimation, change detection, pattern recognition, image processing, computer vision, and data fusion algorithms with improved performance that will significantly impact the effectiveness of DOD in recognizing spatio-temporal patterns of activity in heterogeneous volumes of data. The research will produce much needed models of behavior and will take a stand against the wrongheaded idea that one can infer goals and track behaviors in a purely data-driven way, without rich models. The idea of developing rich models and embedding them in nonlinear estimators and classifiers, and integrating trajectories through multiple, compensating spaces has great potential for higher levels of information fusion, i.e., to the sort of activity conventionally assigned to level 2, 3 and 4 fusion tasks. We believe that our research will result in practical and scalable algorithms for on-line tracking of plans and intentions, which can serve as a basis for an automatic real-time detection and warning system for intelligence applications. Such methodology has direct application to information assurance, MASINT vulnerability assessment, video surveillance and defense against terrorism.

6. FUTURE TECHNOLOGY TRANSFER

The research will produce effective methods and algorithms as well as related software products of interest for target tracking and intelligence analysis communities.

The design of the robotics testbed has led to the formation of a new company, RoboES, from former students at UCLA, that has recently submitted an SBIR proposal. Andrea Bertozzi consults for a defense contractor who is working on a classified DARPA project involving video tracking. Published works from this MURI project could be useful for the DARPA project. Los Alamos National Laboratory took an interest in the testbed design and jointly sponsored a summer project, described above, on dynamic visibility algorithms with real sensors.

REFERENCES

[1] D. Aberdeen, "A (revised) survey of approximate methods for solving POMDP's," *Technical Report*, Dec. 2003, http://users.rsise.anu.edu.au/ daa/papers.html.

[2] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*. Prentice Hall, Englewood Cliffs, 1993.

[3] D. Bertsekas, *Dynamic Programming*. Prentice-Hall, Upper Saddle River, NJ, 1987.

[4] J. von Brecht, S.R. Thiruvenkadam, and T. Chan, "Occlusion tracking with logical models," preprint, 2007.

[5] H.H. Bui, S. Venkatesh, and G. West, "Policy recognition in the abstract hidden Markov model," *Journal of Artificial Intelligence Research*, vol. 17, pp. 451-499, 2002.

[6] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Proc.*, vol. 10, no. 2, p. 266, 2001.

[7] Y.-L. Chuang, Y.R. Huang, M.R. D'Orsogna, and A.L. Bertozzi, "Multi-vehicle flocking: scalability of cooperative control algorithms using pairwise potentials," *The 2007 IEEE International Conference on Robotics and Automation*, 2007 (accepted).

[8] P.R. Cohen and C.T. Morrison, "The Hats Simulator," *Proceedings of the 2004 Winter Simulation Conference*, pp. 849-856, 2004.

[9] J. Cvitanic, R. Liptser, and B. Rozovskii, "A filtering approach to tracking volatility from prices observed at random times," *Annals of Applied Probability*, vol. 16, no. 3, pp. 1633-1652, 2006.

[10] J. Cvitanic, B. Rozovskii, and I. Zalyapin, "Numerical estimation of volatility values from discretely observed diffusion data," *J. Numerical Fin.*, 2007 (accepted).

[11] J. Fuemmeler and V.V. Veeravalli, "Smart sleeping policies for energy efficient tracking in sensor networks," Submitted to the *IEEE Transactions on Signal Processing,* October 2006.

[12] J. Fuemmeler and V.V. Veeravalli, "Smart sleeping strategies for localization and tracking in sensor networks," In: *Proc. 40th Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA, November 2006.

[13] W. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov chain Monte Carlo in Practice*. Chapman and Hall, 1996.

[14] A. Harten, B. Engquist, S. Osher, S.R. Chakravarthy, “Uniformly high order accurate essentially nonoscillatory schemes, III,” *Journal of Computational Physics*, vol. 71, pp. 231-303, 1987.

[15] C.H. Hsieh, Y. Chuang, Y. Huang, K.K. Leung, A.L. Bertozzi, and E. Frazzoli, “An Economical micro-car testbed for validation of cooperative control strategies,” *Proc. of the 2006 American Control Conference, Minneapolis*, MN, June 14-16 2006, pp. 1446-1451.

[16] S. Kent, “On the trial of intrusions into information systems,” *IEEE Spectrum*, vol. 37, Issue 12, pp. 52–56, December 2000.

[17] Y. Landa, R. Tsai, and L.-T. Cheng, “Visibility of point clouds and mapping of unknown environments,” *Advanced Concepts for Intelligent Vision Systems, ACIVS 2006*, Sept 18-21, 2006, University of Antwerp, Belgium (preprint available as UCLA CAM report 06-16).

[18] Y. Landa, D. Galkowski, Y.R. Huang, A. Joshi, C. Lee, K.K. Leung, G. Malla, J. Treanor, V. Voroninski, A.L. Bertozzi, and R. Tsai, “Robotic path planning and visibility with limited sensor data,” *The 2007 American Control Conference*, 2007 (to appear).

[19] K.K. Leung, C.H. Hsieh, Y.R. Huang, A. Joshi, V. Voroninski, and A.L. Bertozzi, “A second generation micro-vehicle testbed for cooperative control and sensing strategies,” *The 2007 American Control Conference*, 2007 (to appear).

[20] M. Moelich and T. Chan, “Joint segmentation and registration using logic models,” *J. Vis. Commun. Image R.*, vol. 15, pp. 333358, 2005.

[21] D. Reid, “An algorithm for tracking multiple targets,” *IEEE Trans. Automat. Contr.*, vol. 24, no. 6, pp. 84-90, 1979.

[22] S. Saria and S. Mahadevan, “Probabilistic plan recognition in multiagent systems,” *Proceedings of ICAPS-04*, 2004.

[23] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York, 1985.

[24] A.G. Tartakovsky, *Sequential Methods in the Theory of Information Systems*. Radio and Communications, Moscow, 1991.

[25] A.G. Tartakovsky, “Asymptotic performance of a multichart CUSUM test under false alarm probability constraint,” *Proc. 44th IEEE Conference on Decision and Control and the European Control Conference (CDC-ECC’05)*, December 12-15, 2005, pp. 320–325, Seville, Spain, Omnipress CD-ROM, ISBN 0-7803-9568-9.

[26] A.G. Tartakovsky and H. Kim, “Performance of certain decentralized distributed change detection procedures,” *Proc. 9th International Conference on Information Fusion*, Florence, Italy, 10-13 July 2006, CD ISBN 0-9721844-6-5, IEEE Catalog No. 06EX1311C.

[27] A.G. Tartakovsky, B.L. Rozovskii, R.B. Blažek, and H. Kim, “Detection of intrusions in information systems by sequential change-point methods (with discussion),” *Statistical Methodology*, vol. 3, no. 3, pp. 252-340, 2006.

[28] A.G. Tartakovsky, B.L. Rozovskii, R.B. Blažek, and H. Kim , “A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods,” *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3372-3382, 2006.

[29] A.G. Tartakovsky, B.L. Rozovskii, and K. Shah, "A nonparametric multichart CUSUM test for rapid intrusion detection," *JSM Proceedings (CD Rom)*, Minneapolis, MN, 2005.

[30] A.G. Tartakovsky and V. Veeravalli, "Change-point detection in multichannel and distributed systems with applications," In: *Applications of Sequential Methodologies* (N. Mukhopadhyay, S. Datta and S. Chattopadhyay, eds.), Marcel Dekker, Inc., New York, pp. 339-370, 2004.

[31] A. Tartakovsky and V. Veeravalli, "An efficient sequential procedure for detecting changes in multichannel and distributed systems," *Proceedings of the 5th International Conference on Information Fusion*, Annapolis, MD, 8-11 July 2002, vol. 1, pp. 41–48.

[32] A. Tartakovsky and V. Veeravalli, "Quickest Change Detection in Distributed Sensor Systems," *Proceedings of the 6th International Conference on Information Fusion*, Australia, 8-11 July 2003.