

The 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 comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 26-01-2017	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 1-Sep-2015 - 31-Aug-2016
---	--------------------------------	--

4. TITLE AND SUBTITLE Final Report: W911NF-12-R-0012-03: Distributed Fusion with Copula: ARO - Research Area 5: Computing Science, in particular, 5.2.2 Data and Information Fusion	5a. CONTRACT NUMBER W911NF-15-1-0409
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102

6. AUTHORS Kuo-chu Chang	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES George Mason University 4400 University Drive, MSN 4C6 Fairfax, VA 22030 -4422	8. PERFORMING ORGANIZATION REPORT NUMBER
--	--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 67776-CS.5

12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited
--

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

14. ABSTRACT The research objective of this STIR effort is to address the fundamental problems of distributed fusion in networks with dynamic topology and heterogeneous sensor characteristics. Emerging information acquisition systems such as cloud robotics require fusion applications for ad hoc and often decentralized collaborations among fusion processes and heterogeneous data sources. Those dynamic environments encompass conditions that are inconsistent with existing data fusion algorithm assumptions. Specifically, the heavy reliance on conditional independence assumptions among sensor observations and state estimates is essential to many traditional fusion algorithms. The
--

15. SUBJECT TERMS distributed fusion, copula

16. SECURITY CLASSIFICATION OF:	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Kuo-Chu Chang
a. REPORT UU	UU		19b. TELEPHONE NUMBER 703-993-1639
b. ABSTRACT UU			
c. THIS PAGE UU			

Report Title

Final Report: W911NF-12-R-0012-03: Distributed Fusion with Copula: ARO - Research Area 5: Computing Science, in particular, 5.2.2 Data and Information Fusion

ABSTRACT

The research objective of this STIR effort is to address the fundamental problems of distributed fusion in networks with dynamic topology and heterogeneous sensor characteristics. Emerging information acquisition systems such as cloud robotics require fusion applications for ad hoc and often decentralized collaborations among fusion processes and heterogeneous data sources. Those dynamic environments encompass conditions that are inconsistent with existing data fusion algorithm assumptions. Specifically, the heavy reliance on conditional independence assumptions among sensor observations and state estimates is essential to many traditional fusion algorithms. The research seeks to develop distributed fusion theories and algorithms that relax the dependence on knowledge about sensing characteristics and network dynamics. Specifically, it investigates copula theory and associated approaches for scalable, multi-modal data fusion methodologies. The goals are to develop reliable fusion theory and algorithms and evaluate them in a realistic simulated environment for performance validation.

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

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

<u>Received</u>	<u>Paper</u>
01/25/2017	1 Kelin Lu, KC Chang, and Rui Zhou. Distributed Fusion with Copula, Journal in Mathematical Problems in Engineering, (): . doi:
TOTAL:	1

Number of Papers published in peer-reviewed journals:

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

<u>Received</u>	<u>Paper</u>
TOTAL:	

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

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

<u>Received</u>	<u>Paper</u>
01/25/2017	2 CY Chong, KC Chang, and Shozo Mori. Comparison of Optimal Distributed Estimation and Consensus Filtering, 19th International Conference on Information Fusion. 05-JUL-16, Heidelberg, Germany. : ,
01/25/2017	3 Sabyasachi Guharay, KC Chang and Jie Xu. Robust Estimation of Value-at-Risk through Correlated Frequency and Severity Model, 19th International Conference on Information Fusion. 05-JUL-16, Heidelberg, Germany. : ,
TOTAL:	2

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

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

<u>Received</u>	<u>Paper</u>
-----------------	--------------

TOTAL:

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

(d) Manuscripts

<u>Received</u>	<u>Paper</u>
-----------------	--------------

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	<u>Discipline</u>
Sabyasachi Guharay	0.20	
Rama Chandra Raju Rajasagi.	0.50	
FTE Equivalent:	0.70	
Total Number:	2	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	National Academy Member
KC Chang	0.25	
FTE Equivalent:	0.25	
Total Number:	1	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Student Metrics

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

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

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

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

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

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

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

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

Names of Personnel receiving masters degrees

<u>NAME</u>
Total Number:

Names of personnel receiving PHDs

<u>NAME</u>
Sabyasachi Guharay
Total Number:

Names of other research staff

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

INTRODUCTION

Current US Department of Defense, public safety, and commercial Cyber Physical Systems (CPS) are increasing in scope and technological complexity. Most CPS rely on multiple types of distributed sensors to monitor complex system operation. A challenge faced by many existing and emerging application domains—operations management, command and control, or vehicle and robotics automation—is the discovery and conversion of data from multiple but dissimilar sensing sources and types into actionable information for timely and reliable data-driven decision making. This is particularly true for next generation CPS where innovative engineering principles and solutions for a variety of applications are needed for sensing, communication, computing and control algorithms integrated with physical domains.

The fundamental objective of sensing in CPS is to understand, analyze, and optimize performance. Existing theories cannot accurately process and fuse correlated data from distributed heterogeneous sources in a timely and optimal manner. In addition, without a rigorous framework to assess the risk associated with a decision arrived from combining disparate data, the traditional paradigm cannot be effectively applied in dynamic CPS that use inference to enable collaboration in complex and often ad hoc ways.

In ad hoc collaborations, sensors and users are interconnected opportunistically, creating the potential to increase capabilities through shared awareness. These ad hoc sensing networks collect disparate and complementary information throughout the area of interest to meet changing information requirements in dynamic situations. Without a centralized data fusion and management process, the combination of previously-processed results with data from other users and new sensors creates unanticipated correlations among the data. These correlations can produce significant estimation bias and overconfidence that lead to incorrect assessments, decisions, or actions and hence increase the risk of failure in the overall system.

RESEARCH CHALLENGES

A sensor network consists of sensors or sensing nodes and processors or processing nodes, which may be co-located with the sensors. The sensors observe the state of a dynamic system and send the measurements to the processors. The processors communicate with other processors and compute state estimates from the available data. Distributed state estimation involves two steps: communication between nodes and fusion of local and non-local data at each node. When measurements are communicated, distributed estimation is straightforward and involves measurement fusion at multiple nodes. When state estimates are communicated, distributed estimation is more complicated because the state estimates to be fused may have dependent estimation errors due to multiple communication channels between the sensors and processors, as well as common process noise of the dynamic system.

Existing fusion theory and algorithms, however cannot effectively work under those complex and non-deterministic conditions, which lead to unanticipated information exchange and correlations. Existing fusion algorithms rely on known correlations (almost always linear) and homogenous probability distribution characteristics among data. Ad hoc sensing networks may result in unknown data correlations caused by dynamic information exchange. Furthermore, networks of sensors with different modalities need to combine observations from sensors having different probability distributions, some of which may be non-Gaussian. For example, the electro-magnetic features from a Radar sensor and the pixel intensities from an infrared sensor for the same target may be correlated and are not likely to have the same statistical distribution. Generalized models do not exist for inter-modal dependence structures and joint probability distribution determination among statistically dependent multimodal observations.

Two state estimate fusion approaches have been proposed to handle the multiple information path problem. Optimal distributed state estimate fusion identifies the redundant information in the state estimates and reconstructs the global state estimate that could be computed if the measurements in the local estimates were communicated. Algorithms include information matrix filter or tracklet fusion, (with augmented states for non-zero process noise), distributed Kalman filter (DKF) and accumulated state density (ASD) fusion. Because of its good performance and minimal communication requirements, information matrix filter or tracklet fusion is popular for hierarchical fusion architectures where identification and removal of the redundant information is relatively straightforward. However, for sensor networks with more complicated communication patterns, keeping track of the information flow is not trivial and approximation fusion techniques are needed.

Covariance-based state estimate fusion approaches model the dependent errors between the local estimates by cross-covariances and use them to compute the optimal combination of the local estimates, which are in general not the global estimates given all the measurements. Algorithms include maximum likelihood, minimum variance, best linear unbiased estimate, and covariance intersection. With the exception of covariance intersection, which computes a very conservative estimate, covariance-based approaches require knowledge of the information paths of the estimates to be fused and communication of the cross-covariances.

Another approach called consensus filtering does not attempt to compute the optimal estimate given all the sensor measurements or a constrained optimal estimate given the estimates to be fused. Instead of explicitly accounting for the dependent information in the multiple information paths, it uses average consensus to compute the estimates at each node. More specifically, each node computes an estimate and broadcasts to its neighbors. Each neighbor then updates its estimates with a consensus algorithm. Under some fairly general conditions, the consensus estimates for all nodes converge asymptotically to a global estimate. Unfortunately, the approach does not have a strong theoretical justification and the resulting global estimates may not be optimal.

On the other hand, copula based approach to account for the dependent information in the multiple information paths has a solid theoretical justification. While there is an extensive literature on applying the copula theory for signal and image

processing, their applications to data fusion problems has mostly been limited to decision fusion and distributed detection using hypothesis testing. Scalable state estimation and fusion in dynamic networks are challenging both in terms of informational and computational burdens. Their solutions are largely absent from core information fusion literature.

Overall, the primary challenges of the research are:

Challenge 1: Observation Dependency - Sensor data exchange in collaborative systems applications introduces dependencies and coupling among observations. The resulting sensor data can incur complex, nonlinear correlation, making quantification a challenging task. Traditional approaches either force conditional independence assumptions or assume linear correlation, which is often an inadequate characterization.

Challenge 2: Disparate Data - Multimodal sensor networks increase information diversity and availability, but heterogeneous data fusion is particularly difficult. Fusion of estimates and sensor measurements having non-Gaussian distributions and no standard multivariate distribution expressions are largely problem-specific or computationally intensive.

Challenge 3: State Estimate Correlation - Exchanging estimates among fusion agents creates data correlations. Even when conditional independence assumptions are imposed (forcefully) on measurements from different sensors, correlations exist among the local estimates due to common prior information originating from past communication events or internal variables. Existing heuristic fusion rules have degraded performance or misleading uncertainty estimates.

TECHNICAL APPROACH

The proposed method is grounded in the marriage of copula theory and Bayesian graphical approaches for fusing data from disparate sources having complex correlation characteristics. The technique requires no a priori knowledge of network connectivity or communications patterns and will function in a mathematically consistent manner while limiting the amount of data communicated between agents.

In particular, the goal is to integrate the following three components of research:

- A mathematical foundation based on copula and Bayesian probabilistic theory for disparate sensor data with unknown dependency and arbitrary marginal statistical distributions;
- A set of practical and autonomous fusion and decision algorithms integrated with a risk-constrained framework for ad hoc networks with heterogeneous and asynchronous sensing;
- A simulation environment to test and evaluate the algorithm performance in the context of ad hoc networked sensing applications in terms of information exchange and processing requirements.

TECHNICAL SIGNIFICANCE

The increase in computing capabilities and proliferation of diverse sensing devices has resulted in a paradigm shift for complex wireless systems. They have also added many new dimensions to the problem, including scalable and efficient fusion of sensory data. To date, information fusion over complex networks has been limited to simple tasks such as spreading a message or reaching consensus on an average scalar or linear function; meanwhile, the advances in multi-sensor fusion systems have been limited for the most part to highly-engineered network architectures and applications. Primary issues preventing greater fusion capabilities include the long-standing simplification on handling data correlations within the sensing network and fusing data containing non-homogeneous probability distribution.

The method builds on copula theory and set-theoretic derivations of information fusion. Copula theory models dependencies (i. e. correlations) between disparate probability distributions, while set-theoretic information fusion enables probabilistic graphical methods and inference for distributed fusion. Recent proliferation of mobile wireless sensor networks (WSNs) and cloud robotics has added many new potential applications and corresponding challenges for scalable distributed fusion of multisource data. The copula method can potentially reveal true nature of dependencies between variables of disparate probability distributions while set-theoretic information fusion enables probabilistic graphical methods and inference for distributed fusion. The synergy of the copula framework and the graphic model provides an opportunity that has largely not been exploited. This research also opens a door to advance our understanding of relationships among network dynamics, scalability, and system performance in a risk-based decision paradigm. This new paradigm is transformative in the sense that it can be applied to other fields such as social, economic, financial, and systems engineering applications where risk-based policy making is desirable based on a robust uncertainty assessment by analyzing large heterogeneous data. For example, financial and banking industries commonly apply a risk metric called the "Value-at-Risk (VaR)", which is typically defined as the 99% confidence level of the maximum amount of potential loss due to financial transactions or business operations within an institution for a future time period. The banking regulator then specifies a capital requirement for the institution based on its VaR and a pre-determined policy to ensure its liquidity and stability.

TECHNICAL ACCOMPLISHMENTS

The objective of this STIR (Short-Term Innovative Research Program) effort is to conduct research on a new paradigm for distributed fusion in dynamic networks with heterogeneous sensing characteristics. The research project was officially started on September 1, 2015 and completed on August 31, 2016. During the performance period, the PI and his research group have developed a mathematical framework based on copula theory for fusion sensor data with unknown dependency. The method involves identifying a suitable copula function in a distributed fusion formulation to model the inter-modal dependency of local

measurements or state estimates arise from a communication pedigree.

While past research shows that traditional fusion approaches produce inconsistent results in the sense that the perceived uncertainty (error covariance) implied by the fusion algorithm could be significantly smaller than the true one, our research results show that in a hierarchical fusion with feedback architecture, the algorithm based on copula fusion produces consistent estimates (in the sense that the perceived uncertainty is close to the true one) by taking into account the unknown dependency [1]. This is critical because the overly optimistic uncertainty assessment from the traditional algorithms could severely mislead the mission critical decision.

Another related approach, called consensus filtering, avoids explicit handling of the multiple information path problem by using a consensus fusion algorithm for distributed state estimation. However, convergence or consensus does not guarantee optimality of the estimates. Our research results show that consensus on information algorithm performs worse than approximate optimal distributed state fusion using hierarchical fusion with feedback and distributed measurement fusion [2]. Furthermore, consensus algorithms also require more communication. These results suggest that consensus filtering may not be a good approach for distributed estimation and further motivate the proposed approach of using copula in this effort.

While the copula framework was initially developed for a particularly communication architecture (hierarchical fusion with feedback), it can readily be applied to other dynamic networks with arbitrary communication patterns. Note that for scalability, we started with an approximate Gaussian copula in the test scenario in which most of the computation can be executed off-line [1].

The proposed research can also be applied to other fields such as social, economic, financial, and systems engineering applications. Our study on airline and auto accident data shows that a much more accurate estimate of a risk metric can be achieved with the copula approach by fusing heavy-tailed non-Gaussian severity and frequency data with unknown but highly nonlinear correlation [3]. The conventional approaches either severely underestimate or overestimate the values by imposing independence assumptions. While these results can be directly applied to distributed sensing applications, they may also have a significant impact to aviation and insurance industries as well as their regulatory policies [4].

COLLABORATIONS AND POTENTIAL TECHNOLOGY TRANSFER

Dynamic spectrum access in an ad hoc network presents challenges and opportunities for information fusion in a highly dynamic environment. The issues discussed above with respect to decentralized fusion on dynamic sensor networks are applicable to dynamic spectrum sharing. We are incorporating some of the research idea in an application for satellite communication with spectrum sharing. For this work, we are collaborating with Dr. Genshe Chen and Dr. Xin Tian of IFT on an AFOSR small business technology transfer effort (STTR): Decision-Making under Uncertainty for Dynamic Spectrum Access Systems in Space Application [5-9].

We are also collaborating with Dr. Sabyasachi Guharay (a statistician from IRS and a PhD at GMU completed in 2016) [4] on applying the copula fusion theory to fuse and analyze data from multiple sources. The issues of unknown data correlation could affect the accuracy of risk assessment in an IRS operating environment. The goal is to gauge and mitigate operational and fraud losses as accurate as possible by fusing large relevant data. Our recent research results shows that a much more accurate risk assessment can be achieved with the copula approach by fusing multiple correlated data with unknown dependency [3-4].

SCIENTIFIC ACCOMPLISHMENTS

This project addresses the challenge of sensing and data fusion in complex and dynamic environments - a key signature in almost all cyber-physical systems. The main aspect of this research is the joint development of theories and algorithms for scalable, decentralized, heterogeneous data fusion techniques combined with a rigorous risk assessment. In particular, it has contributed to the following research areas: (1) a novel mathematical framework based on copula and Bayesian probability theory, (2) a set of autonomous and scalable fusion algorithms for hybrid sensing networks with asynchronous and arbitrary connectivity, (3) a test and simulation environment with realistic sensing networks for comprehensive performance evaluation, and (4) a novel risk-based framework developed for general decision making problems under uncertainty. The effort provides a preliminary quantitative and qualitative understanding of data correlation and scalability, providing a potential to resolve distributed fusion problems that cannot be sufficiently addressed with existing theories and methods.

Publications partially supported by the effort

[1] Kelin Lu, KC Chang, and Rui Zhou, "Distributed Fusion with Copula," Submitted to Journal in Mathematical Problems in Engineering, June, 2016.

[2] CY Chong, KC Chang, and Shozo Mori, "Comparison of Optimal Distributed Estimation and Consensus Filtering," in proc. 19th International Conference on Information Fusion, July 5-8, 2016, Heidelberg, Germany.

[3] Sabyasachi Guharay, KC Chang and Jie Xu, "Robust Estimation of Value-at-Risk through Correlated Frequency and Severity Model," in proc. 19th International Conference on Information Fusion, July 5-8, 2016, Heidelberg, Germany.

[4] Sabyasachi Guharay, "Methodological and Empirical Investigations in Quantification of Modern Operational Risk Management," Ph.D. dissertation, Dept. of Systems Engineering & Operations Research, George Mason University, August, 2016.

[5] Todd Martin and KC Chang, "Probabilistic Reasoning and Risk-Constrained Dynamic Spectrum Access," submitted to IEEE Trans. on Wireless Communications, Oct., 2016.

[6] Todd Martin and KC Chang, "Probabilistic Reasoning for Secondary Spectrum Sharing: A Mechanism for Increasing Spectrum Access Efficiency and Assessing Alternative Secondary Sharing Models," in Proc. TPRC 44, Sept., 2016.

[7] Todd Martin, KC Chang, Xin Tian, Genshe Chen, "Analysis of a Probability-Based SATCOM Situational Awareness Model for Parameter Estimation," SPIE Defense and Security Symposium, Baltimore, Maryland, 2016.

[8] Todd Martin, KC Chang, Xin Tian, and Genshe Chen, "Analysis of a Probability-Based SATCOM Situational Awareness Model for Parameter Estimation," SPIE, April, 2016.

[9] Todd Martin, "DSA", Ph.D. dissertation, Dept. of Systems Engineering & Operations Research, George Mason University, December, 2016.

Technology Transfer

Dynamic spectrum access in an ad hoc network presents challenges and opportunities for information fusion in a highly dynamic environment. The issues investigated in the research with respect to decentralized fusion on dynamic sensor networks are applicable to dynamic spectrum sharing. We are incorporating some of the research idea in an application for satellite communication with spectrum sharing. For this work, we are collaborating with Dr. Genshe Chen and Dr. Xin Tian of IFT on an AFOSR small business technology transfer effort (STTR): Decision-Making under Uncertainty for Dynamic Spectrum Access Systems in Space Application.

We are also collaborating with Dr. Sabyasachi Guharay (a statistician from IRS and a PhD at GMU completed in 2016) on applying the copula fusion theory to fuse and analyze data from multiple sources. The issues of unknown data correlation could affect the accuracy of risk assessment in an IRS operating environment. The goal is to gauge and mitigate operational and fraud losses as accurate as possible by fusing large relevant data. Our recent research results shows that a much more accurate risk assessment can be achieved with the copula approach by fusing multiple correlated data with unknown dependency.

FINAL REPORT

ARO Short Term Innovative Research (STIR)

Grant #W911NF-15-1-0409

Title: Distributed Fusion with Copula

Performance period: Sept. 1, 2015 – August 31, 2016

Prime Offeror: George Mason University

Technical POC: Dr. KC Chang
Systems Engineering and Operations Research Department
George Mason University
4400 University Dr., MS 4A6
Fairfax, VA 22030
Voice : (703) 993-1639
Fax : (703) 993-1521
kchang@gmu.edu

Administrative POC: Donna Senator
Associate Director
George Mason University
4400 University Dr., MS 4A6
Fairfax, VA 22030
Voice : (703) 993-4806
Fax : (703) 993-2296
ospgrant@gmu.edu

ABSTRACT

The research objective of this STIR effort is to address the fundamental problems of distributed fusion in networks with dynamic topology and heterogeneous sensor characteristics. Emerging information acquisition systems such as cloud robotics require fusion applications for ad hoc and often decentralized collaborations among fusion processes and heterogeneous data sources. Those dynamic environments encompass conditions that are inconsistent with existing data fusion algorithm assumptions. Specifically, the heavy reliance on conditional independence assumptions among sensor observations and state estimates is essential to many traditional fusion algorithms. The research seeks to develop distributed fusion theories and algorithms that relax the dependence on knowledge about sensing characteristics and network dynamics. Specifically, it investigates copula theory and associated approaches for scalable, multi-modal data fusion methodologies. The goals are to develop reliable fusion theory and algorithms and evaluate them in a realistic simulated environment for performance validation.

INTRODUCTION

Current US Department of Defense, public safety, and commercial Cyber Physical Systems (CPS) are increasing in scope and technological complexity. Most CPS rely on multiple types of distributed sensors to monitor complex system operation. A challenge faced by many existing and emerging application domains—operations management, command and control, or vehicle and robotics automation—is the discovery and conversion of data from multiple but dissimilar sensing sources and types into actionable information for timely and reliable data-driven decision making. This is particularly true for next generation CPS where innovative engineering principles and solutions for a variety of applications are needed for sensing, communication, computing and control algorithms integrated with physical domains.

The fundamental objective of sensing in CPS is to understand, analyze, and optimize performance. Existing theories cannot accurately process and fuse correlated data from distributed heterogeneous sources in a timely and optimal manner. In addition, without a rigorous framework to assess the risk associated with a decision arrived from combining disparate data, the traditional paradigm cannot be effectively applied in dynamic CPS that use inference to enable collaboration in complex and often ad hoc ways.

In ad hoc collaborations, sensors and users are interconnected opportunistically, creating the potential to increase capabilities through shared awareness. These ad hoc sensing networks collect disparate and complementary information throughout the area of interest to meet changing information requirements in dynamic situations. Without a centralized data fusion and management process, the combination of previously-processed results with data from other users and new sensors creates unanticipated correlations among the data. These correlations can produce significant estimation bias and overconfidence that lead to incorrect assessments, decisions, or actions and hence increase the risk of failure in the overall system.

RESEARCH CHALLENGES

A sensor network consists of sensors or sensing nodes and processors or processing nodes, which may be co-located with the sensors. The sensors observe the state of a dynamic system and send

the measurements to the processors. The processors communicate with other processors and compute state estimates from the available data. Distributed state estimation involves two steps: communication between nodes and fusion of local and non-local data at each node. When measurements are communicated, distributed estimation is straightforward and involves measurement fusion at multiple nodes. When state estimates are communicated, distributed estimation is more complicated because the state estimates to be fused may have dependent estimation errors due to multiple communication channels between the sensors and processors, as well as common process noise of the dynamic system.

Existing fusion theory and algorithms, however cannot effectively work under those complex and non-deterministic conditions, which lead to unanticipated information exchange and correlations. Existing fusion algorithms rely on known correlations (almost always linear) and homogenous probability distribution characteristics among data. Ad hoc sensing networks may result in unknown data correlations caused by dynamic information exchange. Furthermore, networks of sensors with different modalities need to combine observations from sensors having different probability distributions, some of which may be non-Gaussian. For example, the electro-magnetic features from a Radar sensor and the pixel intensities from an infrared sensor for the same target may be correlated and are not likely to have the same statistical distribution. Generalized models do not exist for inter-modal dependence structures and joint probability distribution determination among statistically dependent multimodal observations.

Two state estimate fusion approaches have been proposed to handle the multiple information path problem. Optimal distributed state estimate fusion identifies the redundant information in the state estimates and reconstructs the global state estimate that could be computed if the measurements in the local estimates were communicated. Algorithms include information matrix filter or tracklet fusion, (with augmented states for non-zero process noise), distributed Kalman filter (DKF) and accumulated state density (ASD) fusion. Because of its good performance and minimal communication requirements, information matrix filter or tracklet fusion is popular for hierarchical fusion architectures where identification and removal of the redundant information is relatively straightforward. However, for sensor networks with more complicated communication patterns, keeping track of the information flow is not trivial and approximation fusion techniques are needed.

Covariance-based state estimate fusion approaches model the dependent errors between the local estimates by cross-covariances and use them to compute the optimal combination of the local estimates, which are in general not the global estimates given all the measurements. Algorithms include maximum likelihood, minimum variance, best linear unbiased estimate, and covariance intersection. With the exception of covariance intersection, which computes a very conservative estimate, covariance-based approaches require knowledge of the information paths of the estimates to be fused and communication of the cross-covariances.

Another approach called consensus filtering does not attempt to compute the optimal estimate given all the sensor measurements or a constrained optimal estimate given the estimates to be fused. Instead of explicitly accounting for the dependent information in the multiple information paths, it uses average consensus to compute the estimates at each node. More specifically, each node computes an estimate and broadcasts to its neighbors. Each neighbor then updates its estimates with a consensus algorithm. Under some fairly general conditions, the consensus estimates for all nodes converge asymptotically to a global estimate. Unfortunately, the

approach does not have a strong theoretical justification and the resulting global estimates may not be optimal.

On the other hand, copula based approach to account for the dependent information in the multiple information paths has a solid theoretical justification. While there is an extensive literature on applying the copula theory for signal and image processing, their applications to data fusion problems has mostly been limited to decision fusion and distributed detection using hypothesis testing. Scalable state estimation and fusion in dynamic networks are challenging both in terms of informational and computational burdens. Their solutions are largely absent from core information fusion literature.

Overall, the primary challenges of the research are:

Challenge 1: Observation Dependency - Sensor data exchange in collaborative systems applications introduces dependencies and coupling among observations. The resulting sensor data can incur complex, nonlinear correlation, making quantification a challenging task. Traditional approaches either force conditional independence assumptions or assume linear correlation, which is often an inadequate characterization.

Challenge 2: Disparate Data - Multimodal sensor networks increase information diversity and availability, but heterogeneous data fusion is particularly difficult. Fusion of estimates and sensor measurements having non-Gaussian distributions and no standard multivariate distribution expressions are largely problem-specific or computationally intensive.

Challenge 3: State Estimate Correlation - Exchanging estimates among fusion agents creates data correlations. Even when conditional independence assumptions are imposed (forcefully) on measurements from different sensors, correlations exist among the local estimates due to common prior information originating from past communication events or internal variables. Existing heuristic fusion rules have degraded performance or misleading uncertainty estimates.

TECHNICAL APPROACH

The proposed method is grounded in the marriage of copula theory and Bayesian graphical approaches for fusing data from disparate sources having complex correlation characteristics. The technique requires no *a priori* knowledge of network connectivity or communications patterns and will function in a mathematically consistent manner while limiting the amount of data communicated between agents.

In particular, the goal is to integrate the following three components of research:

- A mathematical foundation based on copula and Bayesian probabilistic theory for disparate sensor data with unknown dependency and arbitrary marginal statistical distributions;
- A set of practical and autonomous fusion and decision algorithms integrated with a risk-constrained framework for ad hoc networks with heterogeneous and asynchronous sensing;
- A simulation environment to test and evaluate the algorithm performance in the context of ad hoc networked sensing applications in terms of information exchange and processing requirements.

TEHNICAL SIGNIFICANCE

The increase in computing capabilities and proliferation of diverse sensing devices has resulted in a paradigm shift for complex wireless systems. They have also added many new dimensions to the problem, including scalable and efficient fusion of sensory data. To date, information fusion over complex networks has been limited to simple tasks such as spreading a message or reaching consensus on an average scalar or linear function; meanwhile, the advances in multi-sensor fusion systems have been limited for the most part to highly-engineered network architectures and applications. Primary issues preventing greater fusion capabilities include the long-standing simplification on handling data correlations within the sensing network and fusing data containing non-homogeneous probability distribution.

The method builds on *copula* theory and set-theoretic derivations of information fusion. Copula theory models dependencies (i.e. correlations) between disparate probability distributions, while set-theoretic information fusion enables probabilistic graphical methods and inference for distributed fusion. Recent proliferation of mobile wireless sensor networks (WSNs) and cloud robotics has added many new potential applications and corresponding challenges for scalable distributed fusion of multisource data. The copula method can potentially reveal true nature of dependencies between variables of disparate probability distributions while set-theoretic information fusion enables probabilistic graphical methods and inference for distributed fusion. The synergy of the copula framework and the graphic model provides an opportunity that has largely not been exploited.

This research also opens a door to advance our understanding of relationships among network dynamics, scalability, and system performance in a risk-based decision paradigm. This new paradigm is transformative in the sense that it can be applied to other fields such as social, economic, financial, and systems engineering applications where risk-based policy making is desirable based on a robust uncertainty assessment by analyzing large heterogeneous data. For example, financial and banking industries commonly apply a risk metric called the “Value-at-Risk (VaR)”, which is typically defined as the 99% confidence level of the maximum amount of potential loss due to financial transactions or business operations within an institution for a future time period. The banking regulator then specifies a capital requirement for the institution based on its VaR and a pre-determined policy to ensure its liquidity and stability.

TECHNICAL ACCOMPLISHMENTS

The objective of this STIR (Short-Term Innovative Research Program) effort is to conduct research on a new paradigm for distributed fusion in dynamic networks with heterogeneous sensing characteristics. The research project was officially started on September 1, 2015 and completed on August 31, 2016. During the performance period, the PI and his research group have developed a mathematical framework based on copula theory for fusion sensor data with unknown dependency. The method involves identifying a suitable copula function in a distributed fusion formulation to model the inter-modal dependency of local measurements or state estimates arise from a communication pedigree.

While past research shows that traditional fusion approaches produce inconsistent results in the sense that the perceived uncertainty (error covariance) implied by the fusion algorithm could be

significantly smaller than the true one, our research results show that in a hierarchical fusion with feedback architecture, the algorithm based on copula fusion produces consistent estimates (in the sense that the perceived uncertainty is close to the true one) by taking into account the unknown dependency [1]. This is critical because the overly optimistic uncertainty assessment from the traditional algorithms could severely mislead the mission critical decision.

Another related approach, called consensus filtering, avoids explicit handling of the multiple information path problem by using a consensus fusion algorithm for distributed state estimation. However, convergence or consensus does not guarantee optimality of the estimates. Our research results show that consensus on information algorithm performs worse than approximate optimal distributed state fusion using hierarchical fusion with feedback and distributed measurement fusion [2]. Furthermore, consensus algorithms also require more communication. These results suggest that consensus filtering may not be a good approach for distributed estimation and further motivate the proposed approach of using copula in this effort.

While the copula framework was initially developed for a particularly communication architecture (hierarchical fusion with feedback), it can readily be applied to other dynamic networks with arbitrary communication patterns. Note that for scalability, we started with an approximate Gaussian copula in the test scenario in which most of the computation can be executed off-line [1].

The proposed research can also be applied to other fields such as social, economic, financial, and systems engineering applications. Our study on airline and auto accident data shows that a much more accurate estimate of a risk metric can be achieved with the copula approach by fusing heavy-tailed non-Gaussian severity and frequency data with unknown but highly nonlinear correlation [3]. The conventional approaches either severely underestimate or overestimate the values by imposing independence assumptions. While these results can be directly applied to distributed sensing applications, they may also have a significant impact to aviation and insurance industries as well as their regulatory policies [4].

COLLABORATIONS AND POTENTIAL TECHNOLOGY TRANSFER

Dynamic spectrum access in an ad hoc network presents challenges and opportunities for information fusion in a highly dynamic environment. The issues discussed above with respect to decentralized fusion on dynamic sensor networks are applicable to dynamic spectrum sharing. We are incorporating some of the research idea in an application for satellite communication with spectrum sharing. For this work, we are collaborating with Dr. Genshe Chen and Dr. Xin Tian of IFT on an AFOSR small business technology transfer effort (STTR): Decision-Making under Uncertainty for Dynamic Spectrum Access Systems in Space Application [5-9].

We are also collaborating with Dr. Sabyasachi Guharay (a statistician from IRS and a PhD at GMU completed in 2016) [4] on applying the copula fusion theory to fuse and analyze data from multiple sources. The issues of unknown data correlation could affect the accuracy of risk assessment in an IRS operating environment. The goal is to gauge and mitigate operational and fraud losses as accurate as possible by fusing large relevant data. Our recent research results shows that a much more accurate risk assessment can be achieved with the copula approach by fusing multiple correlated data with unknown dependency [3-4].

SCIENTIFIC ACCOMPLISHMENTS

This project addresses the challenge of sensing and data fusion in complex and dynamic environments - a key signature in almost all cyber-physical systems. The main aspect of this research is the joint development of theories and algorithms for scalable, decentralized, heterogeneous data fusion techniques combined with a rigorous risk assessment. In particular, it has contributed to the following research areas: (1) a novel mathematical framework based on copula and Bayesian probability theory, (2) a set of autonomous and scalable fusion algorithms for hybrid sensing networks with asynchronous and arbitrary connectivity, (3) a test and simulation environment with realistic sensing networks for comprehensive performance evaluation, and (4) a novel risk-based framework developed for general decision making problems under uncertainty. The effort provides a preliminary quantitative and qualitative understanding of data correlation and scalability, providing a potential to resolve distributed fusion problems that cannot be sufficiently addressed with existing theories and methods.

Publications partially supported by the effort

- [1] Kelin Lu, KC Chang, and Rui Zhou, "Distributed Fusion with Copula," Submitted to Journal in Mathematical Problems in Engineering, June, 2016.
- [2] CY Chong, KC Chang, and Shozo Mori, "Comparison of Optimal Distributed Estimation and Consensus Filtering," in proc. *19th International Conference on Information Fusion*, July 5-8, 2016, Heidelberg, Germany.
- [3] Sabyasachi Guharay, KC Chang and Jie Xu, "Robust Estimation of Value-at-Risk through Correlated Frequency and Severity Model," in proc. *19th International Conference on Information Fusion*, July 5-8, 2016, Heidelberg, Germany.
- [4] Sabyasachi Guharay, "Methodological and Empirical Investigations in Quantification of Modern Operational Risk Management," Ph.D. dissertation, Dept. of Systems Engineering & Operations Research, George Mason University, August, 2016.
- [5] Todd Martin and KC Chang, "Probabilistic Reasoning and Risk-Constrained Dynamic Spectrum Access," submitted to *IEEE Trans. on Wireless Communications*, Oct., 2016.
- [6] Todd Martin and KC Chang, "Probabilistic Reasoning for Secondary Spectrum Sharing: A Mechanism for Increasing Spectrum Access Efficiency and Assessing Alternative Secondary Sharing Models," in Proc. TPRC 44, Sept., 2016.
- [7] Todd Martin, KC Chang, Xin Tian, Genshe Chen, "Analysis of a Probability-Based SATCOM Situational Awareness Model for Parameter Estimation," *SPIE Defense and Security Symposium*, Baltimore, Maryland, 2016.
- [8] Todd Martin, KC Chang, Xin Tian, and Genshe Chen, "Analysis of a Probability-Based SATCOM Situational Awareness Model for Parameter Estimation," *SPIE*, April, 2016.
- [9] Todd Martin, "DSA", Ph.D. dissertation, Dept. of Systems Engineering & Operations Research, George Mason University, December, 2016.