

Data Pedigree and Strategies for Dynamic Level-One Sensor Data Fusion

Marion G. Ceruti

Space and Naval Warfare Systems Center, San
Diego (SSC-SD), Code 246206
53560 Hull Street, San Diego, CA 92152-5001,
marion.ceruti@navy.mil

Brenda J. Powers

Space and Naval Warfare Systems Center, San
Diego (SSC-SD), Code 24226
53560 Hull Street, San Diego, CA 92152-5001,
bpowers@spawar.navy.mil

Tedd L. Wright

Space and Naval Warfare Systems Center, San
Diego (SSC-SD), Code 2725
53560 Hull Street, San Diego, CA 92152-5001
tedd.wright@navy.mil

Scott C. McGirr

Space and Naval Warfare Systems Center, San
Diego (SSC-SD), Code 2725
53560 Hull Street, San Diego, CA 92152-5001,
scott.mcgirr@navy.mil

Abstract - *Knowledge Management for Distributed-Tracking (KMDT) is a U.S. Naval research and development project to improve military-communications and information functions in the battle space. These functions include command, control, data fusion, and decision support. It features a scenario for modeling and simulation that shows how knowledge-management technologies, such as ontologies and intelligent agents can improve battle-space awareness and the decision-making process in command centers with respect to distributed tracking and threat identification of targets. Data on cross lines of bearings can be acquired from sensors using a secure network. These data and their associated pedigree metadata from multiple platforms in the battle space can be fused to reduce the uncertainty in platform detection, localization, classification and identification (level-one data-fusion object refinement). The pedigree metadata can affect how data are used in fusion tasks.*

Keywords- Decision support, distributed sensing, level-one data fusion, pedigree metadata, sensors, tracking

1 Introduction

The Knowledge Management for Distributed Tracking (KMDT) [4] [5] program assembles technologies to assess the information content exchanged via secure network in the battle space. These technology areas include line-of-bearing (LOBs) cross fixing, sensor and pedigree ontologies [2], [7] and intelligent agents [1]. The advantages of using agents to assist with data fusion are as follow. 1) They are designed to reduce the workload of the sensor-

data analysts by finding, screening and retrieving data that have been chosen according to operator-specific criteria. 2) They take advantage of the network-centric architecture in the FORCENet construct.

An integrated ontology can provide to the agents a set of consistent concepts that describe homogeneous and heterogeneous data fusion in general. A database that has been designed according to a detailed ontology also can serve as a guide to selecting the best alternatives in a given situation, according to sensor characteristics, capabilities, performance, environment, and location. Conventional data-fusion technologies, on the other hand were not designed to take advantage of software agents and ontologies.

A goal of KMDT is to enable military analysts, sensor-system operators, and commanders to reduce uncertainty in command and control by better organizing and using the data collected from existing sensors. Providing pedigree metadata can reduce uncertainty about the validity and reliability of the results of a sensor-data fusion task. Specifically, the KMDT program is testing the hypothesis that cross LOBs, whether heterogeneous or homogeneous can be obtained for the identification and classification of unknown targets. In the present state of the KMDT simulation, the following pedigree metadata are collected for each detected LOB: Location of sensor, in terms of latitude and longitude, date time group of detection, sensor identification number, and frequency detected. These metadata support and accompany each LOB measurement.

Using technology developed in the KMDT [4] [5] and other [10] programs, strategies for data fusion are facilitated that previously were inefficient or impossible given the timeframe for analysis in time-critical scenarios [3].

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the 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 a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE JUL 2006		2. REPORT TYPE		3. DATES COVERED 00-00-2006 to 00-00-2006	
4. TITLE AND SUBTITLE Data Pedigree and Strategies for Dynamic Level-One Sensor Data Fusion				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Space and Naval Warfare Systems Center, San Diego, Code 246206, 53560 Hull Street, San Diego, CA, 92152-5001				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES 9th International Conference on Information Fusion, 10-13 July 2006, Florence, Italy. Sponsored by the International Society of Information Fusion (ISIF), Aerospace & Electronic Systems Society (AES), IEEE, ONR, ONR Global, Selex - Sistemi Integrati, Finmeccanica, BAE Systems, TNO, AFOSR's European Office of Aerospace Research and Development, and the NATO Undersea Research Centre.					
14. ABSTRACT see report					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 5	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

New approaches to tracking, fusion and command and control are now possible in this network-centric environment. During their task execution, intelligent agents can access web portals on the network to obtain information relevant to current requirements for sensor-data and pedigree metadata provided that they are aligned to a common frame of reference in time and space. This common reference frame, sometimes call the common-operating picture or the Single Integrated Picture (SIP) is essential to the success of many data-related activities, including LOB cross fixing and any other data-fusion strategy that relies on the SIP.

Multi-sensor data fusion [9] can be used to refine knowledge of the battle space in ways that are not possible with data from a single sensor. Not only can multiple homogeneous sensors track individual platforms, but also multiple sensor types can participate in a level-one data fusion task [1], [9] (e.g. detection, localization, classification, and identification) coordinated by intelligent agents, thus reducing uncertainty in command and intelligence centers. Agent-assisted data fusion can provide the processed data needed for enhanced, distributed heterogeneous level-one data fusion.

The following terminology will be used throughout the paper. “Platform” means an entity, usually mobile, that has a sensor system, such as a friendly ship or aircraft. “Contact” refers to a detected signal originating from a source that has yet to be localized, tracked, classified or identified. “Target” refers to the object of a level-one fusion task, particularly if at least one signal from the object has been identified or localized.

This paper is organized as follows. Section 2 contrasts data collection with data utilization. Section 3 briefly discusses LOB cross fixing. Section 4 describes pedigree metadata to support level-one fusion. Section 5 describes strategies for level-one sensor data fusion. Section 6 presents a discussion of future directions to conclude the paper.

2 Data collection vs. data utilization

Commanders and sensor operators often are overloaded with tasks and uncorrelated information. Conversely, they sometimes have difficulty in obtaining the correct information they need to make timely decisions, so decisions are made using uncertain information. Locally collected data often are lost or wasted if they cannot be correlated in a timely manner with data from remote sensors and observations. Similarly, data from remote sensors that are not controlled locally would not be transmitted efficiently (and sometimes not at all) if no payoff were perceived for their propagation. To respond rapidly, the commander may need the data that neither are available locally nor transmitted from remote sensors.

The majority of sensor data are not used because the methods and speed of data communications, data fusion, and data analysis have not kept pace with the deployment of sensor arrays and the multitude of sensors that collect data continuously. Some information, such as knowledge implicit in imagery, is difficult to analyze as fast as the

data can be collected. The timeframe for using these data is limited. Data and their associated pedigree metadata must be made available when they are needed for making decisions.

The increased knowledge (as opposed to an increase in only unprocessed data) available in the battle space as a result of the network and SIP also can help reduce the tasking of overloaded operators who could otherwise focus on tasks that are not being done due to lower priorities. Information usually is shared among individuals and groups in communities of interest (COI). Members of the COI need to be able to see, access, and understand the data to accelerate decision making. 2) The take advantage of network technology that was not available in the days of platform-centric warfare. and 3) Being in a position to complete temporarily low-priority tasks may provide an advantage in the battle space because task priorities can change and the outcome of one task affects the execution of another. Some tasks cannot be completed in a timely manner without distributed-network technology, for example, LOB cross fixing in which sensors on board different mobile platforms are used.

3 Registration through LOB cross fixing

Sensors deployed on a single platform, such as a ship, can provide LOB information on unknown contacts and potential targets in their vicinity. Cross LOB targeting (i.e. using sensor data from two ships) either is not done or it is limited to homogeneous sensor systems (e.g. all acoustic sensors). Thus, information about multiple LOBs that could localize the position of a target often does not reach a command center in time to support the decision process. Sometimes operators do not know what to do with new data that are not correlated with existing data. Such data fail to reach the threshold of information to support decision confidence.

4 Pedigree metadata for data fusion

Even when sensor data are available, the reliability of the data may be called into question in the absence of pedigree metadata, the use of which is becoming as important as the data themselves. A datum, whether it is a frequency, a pulse repetition rate, or a line of bearing derives its validity from the manner in which it was obtained, the way it was processed, the reliability of the data with which it was fused, and other information regarding its source.

Collecting sensor data and making them available on a network is necessary but insufficient for operational use. Users are demanding a higher level of reliability and trustworthiness of data and the decisions based on them. The demand is growing for precise target location, selection, and improved data fusion to match the precision of today’s weapon systems. Precision ordnance on the wrong target wastes resources. Sufficient metadata must be made available to support the growing demands of users.

Although the reliability of data reflects a second order uncertainty [6] and is somewhat subjective, many factors

contribute to whether or not data should be trusted and in what context [11]. These factors will vary across many dimensions, such as the level of data fusion and the expectation of the users who are in direct support of the decision makers. At level-one data fusion, data aggregates and their pedigree are simple in comparison to the level of data aggregation and processing for levels two and three (situation assessment and threat assessment respectively). Pedigree collection and propagation at level one data fusion can support the pedigree management at higher levels.

Two categories of pedigree data relate to level-one sensor data fusion depending on when they are applied – before fusion begins and after the data have been processed with fusion engines. Before fusion, the pedigree metadata can assist in making the best data selection to be processed with fusion engines. After fusion, the information that describes the fusion process becomes part of the growing body of metadata that accompanies the finished products and allows reassessment.

Many factors affect the validity and reliability of fused sensor data both before and after fusion. Of these many factors and aspects of metadata pedigree that contribute to the determination of the extent to which data can be trusted [11] and subsequently fused (or not included in the fusion process), the following factors emerge as the most important or useful to level-one fusion:

1. Date-time group (DTG) when the signal was collected. For obvious reasons it is most desirable to fuse data collected at the same time. Otherwise the users could see different pictures of the battle space and possibly draw wrong conclusions [5], [11].
2. Latitude and longitude of the collection platform at the time of collection. Like DTG, the exact location of the collection platform can be critical to the “fusability” of data.
3. Maximum range of sensor [5] compared to the distance between sensor and source – Signals degrade as they propagate, especially through a noisy medium. Sensors may not perform well at the extremes of their ranges.
4. Coverage of sensor [5] compared to the angle from which the signal originates. Coverage may not be equally reliable at the various angles even when detections can be made.
5. Location of possible noise sources, if known.
6. Peak frequency and range of noise compared to the peak frequency of the signal. If noise at 10kHz it might not affect a signal with a peak frequency at 500Hz.
7. Sensor type and mode [5]. Have these types of sensors provided reliable data in the past when used in this mode?
8. Sensor resolution capability [5]. This pertains to spatial and frequency resolution. Can the sensor provide range as well as bearing? What is the error associated with the range measurement? Can the sensor be used to discriminate two signals at whose frequency spectra overlap?
9. Environmental factors [5], such as weather, atmospheric conditions, sea state, and bathythermography that could affect signal propagation through the medium.
10. Reputation of the sensors [11] – Are these sensors known to perform well under the collection conditions?
11. Presence or absence of deceptive signals in the frequency range of the signal’s peak frequencies [11].
12. Presence or absence of alternate sources. Are multiple LOBs available? Which ones are more reliable? Some sources may be utilized simply because no other option is available [11].

Some of these factors that contribute to pedigree metadata consist of static data, such as sensor characteristics and performance, whereas other data are dynamic, such as platform positions and DTGs. When these factors are taken into account in the level-one fusion process, it could affect the fusion strategy for processing recent data and also the collection strategy for future data.

After the fusion process, which can involve various techniques and engines, the pedigree metadata described above still can indicate how much the finished fusion product can be trusted. In addition to these metadata, a complete set of pedigree metadata describing the finished fusion product will include:

1. Names and origins of fusion engines or algorithms.
2. Sensor data or data sets that the engines processed and why they were chosen.
3. The order of processing, if more than one engine or two data were used. (See, for examples [3]).
4. Known limitations of the selected methods.
5. Whether or not the fusion engines were used according to the way in which they were designed to be used. Sometimes a fusion engine with known limitations may be used in the absence of better alternatives.
6. Confidence of conclusion reached and alternate possible hypothesis.

5 Strategies for level-one sensor data fusion

Networks can enable and facilitate knowledge discovery [8] at a level that is not possible in a static, stand-alone environment. Knowing the pedigree of information collected from one source can influence the strategy for collecting other sensor data in the same environment [10]. For example, network communications can help simplify heterogeneous sensor fusion to homogeneous sensor fusion. An example of this is explained below.

Homogeneous sensor fusion has a degree of simplicity that heterogeneous sensor fusion lacks. In homogeneous sensor fusion, the frequency spectra of acoustic signals, for example, can be compared directly. However, in heterogeneous sensor fusion, the spectra of signals obtained from different sensor types (e.g. acoustic and electromagnetic) are not expected to look alike and the con-

clusion that heterogeneous sensors have detected the same contact must be reached through inference. For example one would need to consider the classes of targets that could give rise to both sets of signals.

Using software agents-based technologies, sensor-fusion strategies can be used that previously were either notional only or too difficult or detailed to complete in real time. For example, consider the following scenario. Using an acoustic sensor, Ship A obtains a LOB of an unknown contact. Using agents deployed on the network, an operator on Ship A obtains a cross LOB from Ship B, which obtained the LOB using an electromagnetic sensor.

The operator on Ship A would like to know if the contact detected by Ship B is the same as the contact detected by Ship A. However, even with complete pedigree metadata, not enough information is available for a definitive determination because the sensor types are different. Therefore, the operator on Ship A tries to detect the unknown target using the same sensor type as the one that generated the cross LOB on Ship B. (The updated position can be estimated using dead reckoning.) Now sensor data can be used more efficiently because the signals are directly comparable.

The question arises, why did the operators on ships A and B not use the same sensor type to detect the target during the initial LOB and cross LOB signal detection? Various reasons could include the following:

1. Until the cross LOB was received, the operator on Ship A would not have known which sensor on Ship B had detected the unknown contact.
2. The unknown contact may have been out of the range of the same sensor type at the time of initial detection.
3. A sensor of the same type may not have been operable at the time of the initial detection.

Another strategy is for the sensor analyst on ship B to search for the target using the same sensor type as the one that made the initial contact on Ship A. A comparison of the spectra of the signals would provide a confirmation that the sensors on the two ships had detected the same target. Hence, heterogeneous sensors can lead to homogeneous data fusion. This strategy is expected to work for the detection, localization, tracking, classification, and identification of slow-moving ships. It would have limited utility for the detection of aircraft because the aircraft speeds would preclude a detailed analysis.

One of the advantages of a network-centric architecture [1] is that a commander can task a sensor or request tasking from a sensor that resides on different platforms. This is another way in which web-enabled communications can support data fusion. For example, consider an operator who has three LOBs that are suspected to pertain to the same contact. If these LOBs do not all cross on one point but instead form a triangle, the operator will need to determine which LOBs to use or how to fuse the information provided by all three LOBs. The operator can survey the SIP online and identify a ship that is closer to the contact. The ship's commander can request an additional LOB from a sensor on the remote ship to clarify the position of the contact.

6 Directions for future work

Automated tools are needed to manage pedigree metadata for data analysts who are working on level-one fusion tasks. These software tools are needed to support various strategies described here for level-one data fusion. These tools can include displays that highlight the location of groups of sensors with the same capabilities and their locations in the battle space. Thus, an operator or commander can select the sensor to task for cross LOB or to obtain a report from a homogeneous sensor.

Posting of data and pedigree metadata to shared spaces for COI access is needed except when limited by security or policy constraints.

Public Key Infrastructure (PKI) certificate authentication can improve trust and reliability of sensor data obtained on the network on a node-by-node basis. Experiments and quality assurance testing should be conducted to validate overall network reliability as well as the reliability of sensors and the data they provide to net-based users. Lastly, tagging of data with pedigree metadata can enable knowledge discovery by unanticipated users.

Acknowledgements

The authors thank the Office of Naval Research and the Space and Naval Warfare Systems Center, San Diego, Science and Technology Initiative for their support of this work. This paper is the work of U.S. Government employees performed in the course of employment and no copyright subsists therein. It is approved for public release; distribution is unlimited.

References

- [1] M.G. Ceruti, "Mobile Agents in Network-Centric Warfare," *Institute of Electronics Information and Communication Engineers Transactions on Communications*, Tokyo, Japan, E84-B(10): 2781-2785, Oct. 2001.
- [2] M.G. Ceruti, "Ontology for Level-One Sensor Fusion and Knowledge Discovery," *Proc. of the 2004 Intl. Knowledge Discovery and Ontology Workshop (KDO-2004)*, Sep. 2004.
- [3] M.G. Ceruti and J. Kaina, "Enhancing Dependability of the Battlefield Single Integrated Picture Through Metrics for Modeling and Simulation of Time-Critical Scenarios," *Ninth International IEEE Workshop on Object-oriented Real-time Dependable Systems*, (WORDS 2003F), pp. 69-76, 1-3 Oct. 2003.
- [4] M.G. Ceruti and T.L. Wright, "Knowledge Management for Distributed Tracking and the Next-Generation Command and Control," *IEEE International Software Metrics Symposium (METRICS-2005)*, Sep. 2005.
- [5] M.G. Ceruti, T.L. Wright, D.R. Wilcox, and S.C. McGirr, "Modeling and Simulation for the Knowledge Management for Distributed Tracking (KMDT) Program," *Proceedings of the International Workshop on Modeling and Applied Simulation (MAS-2005)*, pp. 67-75 Oct. 2005.

- [6] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz and F. White, "Revisiting the JDL Data Fusion Model II," *The 7th International Conference on Information Fusion (FUSION 2004)*, June 2004.
- [7] C.J. Matheus, D. Tribble, M.M. Kokar, M.G. Ceruti and Scott C. McGirr, "Towards a Formal Pedigree Ontology for Level-One Sensor Fusion," *Proceedings of the 10th International Command and Control Research and Technology Symposium (ICCRTS 2005)*, June 2005.
- [8] S.C. McGirr, "Knowledge Discovery in a Network," *Proceedings of the National Symposium on Sensors and Data Fusion (NSSDF)*, May 2005.
- [9] E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Boston, 1990.
- [10] J. Waters, M. Stelmach and M.G. Ceruti, "Spiral Systems Implementation Methodology: Application of the Knowledge Web and Network-Centric Best Practices," *World Science and Engineering Academy and Society Transactions on Information Science and Applications*, 2(12): 2088 – 2095, Dec. 2005.
- [11] Y. Gil and D. Artz, "Toward Content Trust of Web Resources," Submitted to the WWW 2006 Conference, May 2006.

Dr. Marion G. Ceruti is a senior scientist in the Command and Control Department at SSC-SD. Dr. Ceruti's professional activities include information-systems research and analysis for command and control decision-support systems, sensor fusion, and research management. She is the author of 90 publications. Dr. Ceruti is a senior member of the IEEE and a member of the IEEE Computer Society, the Association for Computing Machinery, the Acoustical Society of America, the Armed Forces Communications and Electronics Association the International Society for Computers and Their Applications, and the New York Academy of Sciences.

Mr. Tedd L. Wright, is a senior scientist in the Intelligence, Surveillance, and Reconnaissance Department at SSC-SD. He received the BS in mathematics in 1971 and the MS in oceanography in 1976 both from Oregon State University. His areas of expertise are applied mathematics and statistics, physical oceanography, underwater acoustics and optics, modeling and simulation, surveillance systems, data fusion and control, artificial intelligence and autonomous Systems. Mr. Wright has assisted the SSC-SD Corporate Initiatives Group in defining and demonstrating a C4ISR vision and technology roadmap.

Ms. Brenda J. Powers is a scientist in the Command and Control Department at the Space and Naval Warfare Systems Center, San Diego. She received the BS degree from the Point Loma University in 1984 and the MSSE degree from the Naval Postgraduate School in Monterey, CA in the area of collaborative interoperability for decision support. Her professional interests include research in intelligent agent-oriented methodologies, ontologies, and object-oriented analysis and design. She is a member of the ACM and Artificial Intelligence Special Interest Group.

Dr. Scott C. McGirr, is a senior scientist in the Intelligence, Surveillance, and Reconnaissance Department at SSC-SD. He is a department representative on the Science and Technology Advisory Committee (STAC), which oversees discretionary research. He has been a leader in technical areas such as data fusion, systems engineering and earthquake prediction. He has published numerous papers in data fusion and ocean surveillance and has been on the organizing committees of many conferences, including but not limited to the National Symposium on Sensors and Data Fusion (NSSDF).