FEB 17 1997

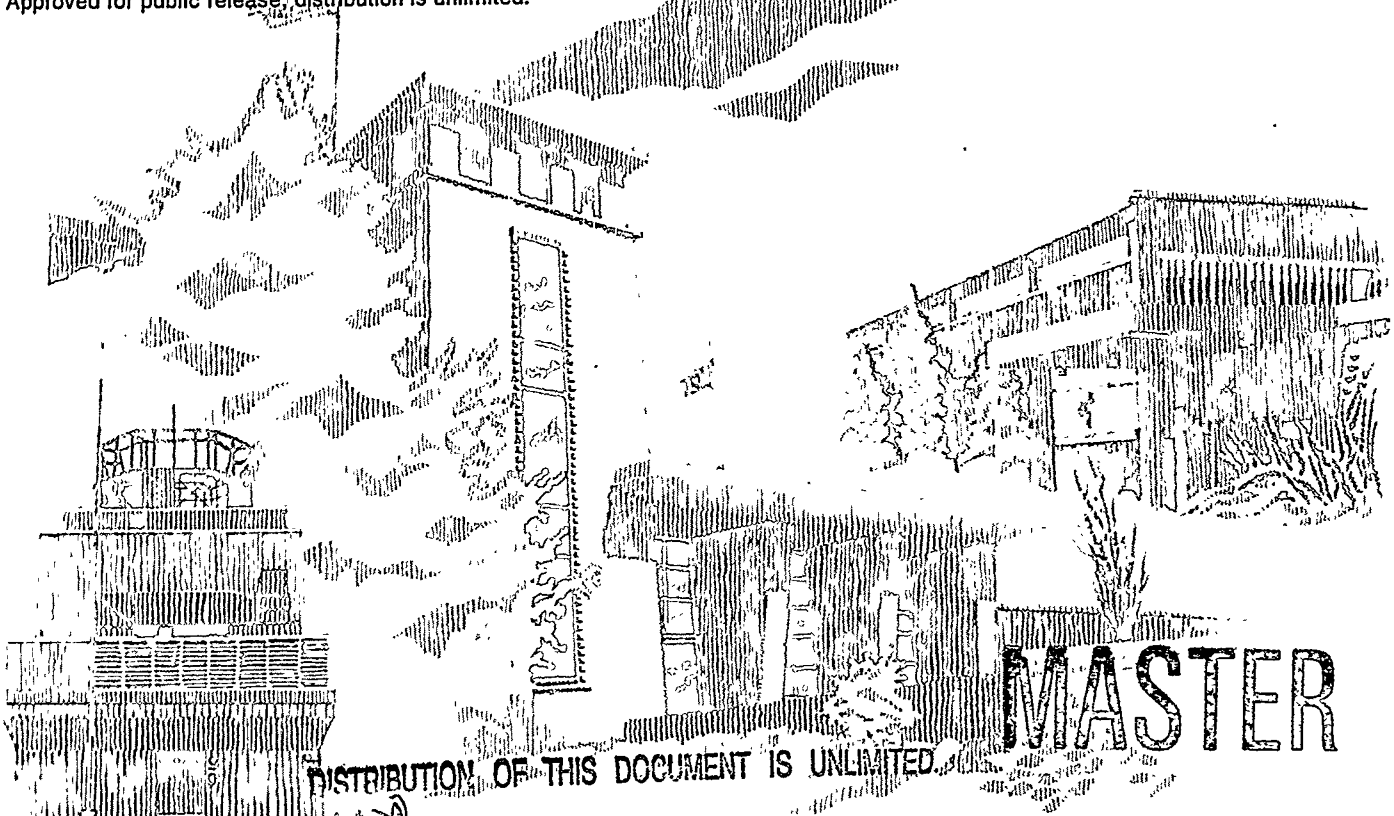
SANDIA REPORT

SAND97–0195 • UC–700
Unlimited Release
Printed January 1997

Information Integration for Data Fusion

Olin H. Bray

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-94AL85000



SF2900Q(8-81)

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A04
Microfiche copy: A01

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Date: February 1, 1997

To: Distribution

From: Olin Bray, 4524

Subject: Final Report for FY93 LDRD
Information Integration for Data Fusion

The attached report is one of three reports that resulted from work done under the FY93 LDRD, *Information Integration for Data Fusion.* Copies are being sent to people who were involved in the project or who might be interested in its results. If you know of other people who would be interested in copies of these reports, please have them contact me or let me know and I will send them a copy.

Purpose of this LDRD:

Data fusion is the integration and analysis of data from multiple sensors to develop a more accurate understanding of a situation and determine how to respond to it. It can be applied in many application areas, several of which were explored in this LDRD project.

The *Information Integration for Data Fusion* LDRD project had two purposes: (1) to see if a natural language-based information modeling methodology could be used for data fusion problems, and if so, (2) to determine whether this methodology would help identify commonalities across areas and achieve greater synergy. Both of these hypotheses were confirmed. The project found five common objects that are the basis for all of the data fusion areas examined: targets, behaviors, environments, signatures, and sensors. Many of the specific facts related to these objects were common across several models and could easily be reused. In some cases, even the terminology remained the same. In other cases, different areas had their own terminology (e.g., a target in defense, a workpiece or machine tool in manufacturing, or an organ for health care), but the concepts were the same. This commonality is important with the growing use of multisensor data fusion. Data fusion is much more difficult if each type of sensor uses its own objects and models rather than building on a common set. Information model integration at the conceptual level is much easier than at the implementation level.

Report 1:

The first report, *Information Integration for Data Fusion* (SAND97-0195) provides a framework for considering data fusion from an information integration perspective, discusses how the synergy generated by this LDRD would have benefited an earlier successful project and contains a summary information model from that project, describes a preliminary truce management information model, and explains how information integration can facilitate cross-treaty synergy for various arms control treaties.

Report 2:

The second report, *Information Model for On-Site Inspection System* (SAND97-0049), describes the information model that was jointly developed as part of two LDRDs: (1) *Information Integration for Data Fusion,* and (2) *Interactive On-Site Inspection System: An Information System to Support Arms Control Inspections.* Section 1 describes the purpose and scope of the two LDRD projects and reviews the prototype development approach, including the use of a GIS. Section 2 describes the information modeling methodology. Section 3 provides a conceptual data dictionary for the OSIS (On-Site Inspection System) model, which can be used in conjunction with the detailed information model provided in the Appendix. Section 4 discussions the lessons learned from the modeling and the prototype. Section 5 identifies the next steps — two alternate paths for future development. The long-term purpose of the On-Site Inspection LDRD was to show the benefits of an information system to support a wide range of on-site inspection activities for both offensive and defensive inspections. The database structure and the information system would support inspection activities under nuclear, chemical, biological, and conventional arms control treaties. This would allow a common database to be shared for all types of inspections, providing much greater cross-treaty synergy. The details of the prototype are described in another Sandia report (SAND93-2300), *Interactive On-Site Inspection System: An Information System to Support Arms Control Inspections.*

Report 3:

The third report, *Data Fusion for Adaptive Control in Manufacturing: Impact on Engineering Information Models* (SAND97-0048), consists of four parts: Section 1 defines data fusion and explains its impact on manufacturing. Section 2 describes an information system architecture and explains the natural language-based information modeling methodology used by this research project. Section 3 identifies the major design and manufacturing functions, reviews the information models required to support them, and then shows how these models must be extended to support data fusion. Section 4 discusses the future directions of this work.

Outside Exposure:

This LDRD work also had exposure outside of Sandia. The first report provided the basis for a presentation, *Information Modeling Framework for Data Fusion Problems,* at the New Mexico DECUS Conference in Albuquerque, NM, in May of 1993. Part of the first report also provided the basis for a panel discussion at the DOE Expo 93 on Intelligence and Special Operations in Oak Ridge, TN. The third report was the basis for a paper, *Data Fusion for Adaptive Control in Manufacturing: Impact on Engineering Information Models,* for the ASME Engineering Information Management Symposium in San Diego in August 1993, which was reprinted in the ASME journal *Computers in Engineering.*

The work that resulted in the second report *(Information Model for On-Site Inspection System)* was done in conjunction with another LDRD that actually developed a prototype system based on the model, which was subsequently demonstrated to IAEA and other agencies. This system is now being shown at the Cooperative Monitoring Center.

SAND97-0195
Unlimited Release
Printed January 1997

Distribution
Category UC-700

Information Integration for Data Fusion

Olin H. Bray
Strategic Business Development
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1378

Abstract

Data fusion has been identified by the Department of Defense as a critical technology for the U.S. defense industry. Data fusion requires combining expertise in two areas — sensors and information integration. Although data fusion is a rapidly growing area, there is little synergy and use of common, reusable, and/or tailorable objects and models, especially across different disciplines, e.g., defense and manufacturing. The Laboratory-Directed Research and Development (LDRD) project had two purposes: (1) to see if a natural language-based information modeling methodology could be used for data fusion problems, and if so, (2) to determine whether this methodology would help identify commonalities across areas and achieve greater synergy. The project confirmed both of the initial hypotheses: (1) that the natural language-based information modeling methodology could be used effectively in data fusion areas and (2) that commonalities could be found that would allow synergy across various data fusion areas. The project found five common objects that are the basis for all of the data fusion areas examined: targets, behaviors, environments, signatures, and sensors. Many of the objects and the specific facts related to these objects were common across several areas and could easily be reused. In some cases, even the terminology remained

(Abstract continued to page ii)

(Abstract continued from title page)

the same. In other cases, different areas had their own terminology (e.g., target in defense, workpiece or machine tool in manufacturing, or an organ in health care), but the concepts were the same. This commonality is important with the growing use of multisensor data fusion. Data fusion is much more difficult if each type of sensor uses its own objects and models rather than building on a common set. Information model integration at the conceptual level is much easier than at the implementation level. Overall, the LDRD project confirmed the benefits of the modeling methodology and found the types of commonality needed to provide synergy across data fusion areas. It also developed some high-level, preliminary information models that can be used as starting points for future data fusion work. This report introduces data fusion, discusses how the synergy generated by this LDRD would have benefited an earlier successful project and contains a summary information model from that project, describes a preliminary truce management information model, and explains how information integration can facilitate cross-treaty synergy for various arms control treaties.

Contents

Figures

Information Integration for Data Fusion

Introduction

Purpose of LDRD

The purpose of this Laboratory-Directed Research and Development (LDRD) was to explore the value of information modeling for data fusion. The project:

- Investigated the application of natural language-based information modeling for a variety of data fusion areas, including defense, arms control, and manufacturing.

- Identified a set of objects — targets, behaviors, environments, signatures, and sensors — that are common to, and the basis for, all of the examined data fusion areas. Many of the specific facts related to these objects were common across several models and could easily be reused. In some cases, even the terminology remained the same. In other cases, different areas had their own terminology, but the concepts were the same. For example, the terms "targets" and "signatures" are not common in manufacturing and health care but many of the facts are the same, so the information models are easily reusable.

- Developed initial high-level information models for several areas and identified their commonality. Information models developed, refined, revisited, or discussed as part of this LDRD include truce management, Automatic Target Recognition/Synthetic Aperture Radar (ATR/SAR), arms control synergy, and adaptive control for manufacturing. The project looked for commonality across the models.

  Jointly with another LDRD, this project also developed an information model to support on-site inspections, which was prototyped as part of the other LDRD. *Information Model for On-Site Inspection System* (SAND97-0049) describes the On-Site Inspection Information Model in detail. Another report, *Data Fusion for Adaptive Control in Manufacturing: Impact on Engineering Information Models* (SAND97-0048) describes the LDRD work and information model for adaptive control for manufacturing. A summary of this work and of the information model was presented at the American Society of

Mechanical Engineers (ASME) Symposium on Engineering Data Management and published in the *International Journal for Engineering with Computers.*

Information modeling uses a natural language, fact-based approach to explicitly model the information requirements for an application area or to integrate a set of application areas. It has been used for requirements definition and information systems development in a variety of areas, but until now not in the data fusion area. This LDRD work showed the value of this approach for defining and designing information systems to support data fusion efforts in a wide range of areas.

Although data fusion is a rapidly growing area, there is little synergy and use of common, reusable, and/or tailorable objects and models, especially across different disciplines, e.g., defense and manufacturing. The project had two purposes: (1) to see if a natural language-based information modeling methodology could be used for data fusion problems and (2) if so, to determine whether this methodology would help identify commonalities across areas and achieve greater synergy.

The project confirmed both of the initial hypotheses: (1) that the natural language-based information modeling methodology could be used effectively in data fusion areas and (2) that commonalities could be found that would allow synergy across various data fusion areas. This commonality is important with the growing use of multisensor data fusion. Data fusion is much more difficult if each type of sensor uses its own objects and models rather than building on a common set. Information model integration at the conceptual level is much easier than at the implementation level.

All of the parts of the report that address information models for specific areas show the benefits of the natural language-based methodology and the commonality and reusability it allows. Section 2, "Data Fusion Lessons for an ATR/SAR Project," also shows the ease with which these information models can be extended for future enhancement. The initial SAR model, which was done before the LDRD, resulted in 450 facts, 70 Fifth Normal Form tables, and an implemented database with 55 tables. At that time, the estimates were that 80 to 90 percent of the model could be reused for another sensor type. As part of the LDRD, this model was reviewed in terms of the common objects and its level of reusability increased. Also we identified some extensions to the original model, which did not include target behavior or multiple sensors. The natural language extensions were easy to do and the extensions were also highly reusable.

Overall the LDRD project confirmed the benefits of the modeling methodology and found the types of commonality needed to provide synergy across data fusion areas. It also developed some high-level, preliminary information models that can be used as starting points for future data fusion work.

Data fusion has been identified by the Department of Defense as a critical technology for the U.S. defense industry. Data fusion requires combining expertise in two areas — sensors and information integration.

This approach to integration for data fusion can create many new types of information-intensive products that Sandia could develop. It would also dramatically improve Sandia's capabilities in information integration.

Organization of the Report

The documentation for this LDRD project consists of three SAND reports: This report is *Information Integration for Data Fusion* (SAND97-0195). The second report, *Information Model for On-Site Inspection System* (SAND97-0049), describes the On-Site Inspection System (OSIS) Model. The OSIS model was done jointly with another LDRD ("Development and Demonstration of an Information System for Arms Control Monitoring and Verification"), described in SAND93-2300, *Interactive On-Site Inspection System: An Information System to Support Arms Control Inspections*. The third report, *Data Fusion for Adaptive Control in Manufacturing: Impact on Engineering Information Models* (SAND97-0048), applies this approach to develop a generic information model for data fusion in a manufacturing environment.

This report consists of six sections:

- Section 1 provides an introduction to data fusion and lays out a framework for understanding and relating the key issues.
- Section 2 reviews an earlier Automatic Target Recognition/Synthetic Aperture Radar (ATR/SAR) project and discusses how the synergy generated by this LDRD would have benefited the project.
- Section 3 provides another example of the use of the methodology, using a preliminary truce management information model.
- Section 4 describes how information integration can facilitate cross-treaty synergy for various arms control treaties.
- Section 5 provides a summary and conclusions for the project.
- Section 6 lists references.

Intentionally Left Blank

1. Data Fusion Introduction and Framework: What and Why Data Fusion?

Data fusion is essentially an information integration problem. It integrates data from multiple sensors to provide better analysis and decision making in a situation than can be done using any single sensor. These sensors may be of the same or different types. Examples of sensor types include radar, thermal, acoustic, laser, optical, and spectrographic analysis or chemical detection devices. Different types of sensors have different strengths and weaknesses. Therefore, integrating data from multiple sensors of different types provides a better result because the strengths of one type can compensate for the weaknesses of another type.

On the battlefield, data fusion may involve integrating data from several radars, IR, and acoustic sensors to better define a target and its characteristics. For arms control and verification, data fusion could involve integrating data from multispectral satellite images, ground-based sensors, and on-site inspections for better analysis. In fact, during the course of the LDRD, the benefits of synergy, both for verification methods within a single arms control treaty and across treaties, became apparent in the modeling and to the policy community.

In addition to these defense applications, advances in data fusion would also benefit many other important areas such as manufacturing (for robotic and adaptive process control), health care (for CAT scans and other types of medical imaging), environment (for the identification and tracking of pollutants), information analysis for intelligence agencies, and transportation (for intelligent traffic control). Therefore, breakthroughs in data fusion will impact many areas of interest to Sandia.

The critical data fusion problem is not data collection and analysis of raw sensor data using complex mathematical algorithms and parallel processors, although these technologies are part of the total solution. The key issue for data fusion is how to convert the initially processed sensor data into information and knowledge to support the decision maker in a timely fashion.

In this LDRD project, the data fusion problem was investigated as an information integration problem. This involved developing a semantic model of several generic scenarios (e.g., on-site inspection, truce management, and adaptive manufacturing). Although some work was done at a detailed level, the real payoff occurred when we generalized the information models to a more generic level, i.e., targets, features, behaviors, environments, signatures, and sensors. Section 2 shows how this generic model enriches and extends the more detailed ATR model and makes its reuse easier and more effective.

What is Data Fusion?

Information systems collect data, process it, provide output, and store data for future use. Traditional information systems get their input from a variety of standard input and use a variety of standard output devices. Sensor-based systems (of which data fusion is a special case) use, at least in part, more specialized input devices called sensors and perhaps specialized output devices called effectors, which cause some action to be taken. For example, sensors may detect an aircraft's position, velocity, and orientation; the system (autopilot) would determine any deviation from the desired flight path; and output signals would be sent to the appropriate effectors to move the control surfaces.

Some systems may take action based on a single reading from a single sensor. Others may base their actions on multiple readings by a single sensor. Still other systems may base their action on one or more readings by multiple sensors. These multiple sensors may be of the same or different types. Sensors may provide a single binary reading (e.g., a door is open or closed), a single continuous reading (e.g., temperature, pressure, or acceleration), or a set of related readings (e.g., multiple readings at a point in time from a specific location, or an image).

Single Sensors

In many systems, readings or images from a single sensor are analyzed to identify a target or an object of interest. The image does not tell you directly whether or not a target is present. It simply provides a signature that must be analyzed to see if it matches the known signature for a specific type of target. If there is a good match, you can infer that the target is present. If there is no match, then you can infer that the target is not present. (In reality the match involves more than just the type of target. It may also involve the target's behavior and the environment in which it is operating.) However, the decision is not really binary. The closer the match, the greater the probability that there is a target present. This is the point at which data fusion, as opposed to reading from just one sensor, becomes important.

Multiple Signatures

Most targets exhibiting a behavior generate multiple types of signatures that can be detected by different types of sensors. For example, a moving tank generates an optical image, a thermal image for an infrared sensor, an acoustic signature, and a seismic signature. If the appropriate sensors were available, these signatures could be captured and analyzed to determine the type of target and behavior. The same tank moving in either a desert or a forest would generate the same signature types, but the actual signatures or patterns would be different. By integrating or fusing data from multiple types of sensors you can balance the weaknesses of some types with the strengths of others. For example, some sensors may be good at detecting precise edges. Others may be less precise with edges but can accurately sense motion. Some may detect minor temperature differences. Other types of sensors detect chemical signatures in the environment such as air or water pollution or the composition of industrial waste streams.

Often a large target such as a tank is decomposed into many distinct features such as turret, gun, treads, and engine compartment. Each of these features provides a distinctive signature for various types of sensors. Data fusion involves processing and integrating all of these signatures to infer what the target is. This can improve both the accuracy (it is a tank) and the precision (it is a T72), and perhaps allow the decision to be determined faster.

Types of Sensors

Multiple sensors may consist of different types of sensors (visual, SAR, IR, acoustic, chemical, etc.) or different sensors of the same type but at different locations, or may be the same sensor but over time. Multiple sensors of the same type can improve your coverage and give you a broader picture of what is happening. Multiple sensors of different types can provide more accurate and complete information.

The strengths of some sensors types can compensate for the weaknesses of others. For example, some sensor types are very good at detecting edges, but are weak at distinguishing motion. Other types of sensors have the opposite characteristics. Still other sensors may be good at distinguishing temperatures, but poor at edges or motion. However, a clear understanding of the complete situation may require all of these types of data, so a suite of sensors would be far better than any individual sensor, no matter where it was placed.

There is often a difference between the signal directly output from the sensor and the variable that is actually read into the system. For example, the pressure on a sensor may be converted into an electronic signal and smoothed out before it is actually passed to a processing system.

A sensor detects some physical phenomenon such as temperature, pressure, light/energy, acceleration, etc., and converts it into a signal that can be displayed or transmitted elsewhere. An infrared sensor does not tell you that the object it sees is a truck or a person. It simply tells you that for a volume in space these are the temperature readings. Additional data is needed to infer what type of object the sensor is actually seeing. In principle, the sensor, the processor, and storage can be packaged into the same box, but there are still three distinct functions. The trend toward "intelligent sensors" involves adding limited additional processing power to the sensors. Examples of the functions included in these intelligent sensors include diagnostics, calibration, and programmability (e.g., on/off and level of sensitivity).

Three Levels of Intelligence Sensors

Nello Zuech (editor of *Handbook of Intelligent Sensors for Industrial Automation*, Addison-Wesley Publishing Co., Reading, MA, 1992) describes three levels of intelligence sensors, although his descriptions relate specifically to industrial automation as opposed to data fusion in a more general sense.

- With Level 1 sensors, the actions are fixed. The sensors initiate a fixed sequence of actions and an operator must handle any exceptions. Level 1 sensors do no data collection.
- Level 2 sensors can adapt their actions based on their input signals. They can also collect and forward data. Level 2 sensors can control "islands of automation."
- Level 3 sensors integrate these islands of automation and adjust the work flow to the islands. They also do data acquisition and are often integrated into plantwide systems.

Data Fusion Applications

Conceptually, data fusion is a class of problems appropriate for many application domains. For defense, it is applicable for a variety of RSTAKA (Reconnaissance, Surveillance, Target Acquisition, and Kill Assessment) functions where data from a variety of sensors must be integrated to deal effectively with many targets of different types.

In manufacturing, multiple sensors must be integrated to control machine tools, work centers, production lines, and factories. This fusion becomes even more important as we move further into computer integrated and agile manufacturing.

Environmental monitoring can also involve integrated data from a variety of sources to get a better understanding of levels and flow of contaminants in air or water supplies (both surface and ground water).

Integrating data from various medical sensors and devices is also an example of data fusion, for either diagnostic purposes or patient monitoring in an Intensive Care Unit, a long-term care facility, or at home.

Finally, transportation work with intelligent vehicles and highways will require extensive use of sensors and data fusion.

Three Types of Environments

In general, sensors and data fusion can be used in three types of environments, often referred to as a designed world, a real world, and a hostile world.

Designed World. In a designed world, such as a machine tool or work center, we have a relatively complete understanding of what exists in that world and how it operates. Furthermore, we have almost complete control of the environment and what goes on in that world.

Real World. A real world, such as with environmental monitoring, is much more complex. We only partially understand the physical phenomena that are being monitored and often have very little control over it. Although our actions may affect this real world, we do not really control it.

Hostile World. Finally, defense applications typically deal with a hostile world. The hostile world is in some ways similar to both the designed and the real worlds. Some parts of it we understand and control, but other parts are less understood and controllable. However, the key factor is that in the hostile world there is usually a hostile opponent who is deliberately responding to our actions in unfriendly ways. In this hostile world, an opponent may try to jam the sensors and create false readings so that targets are not identified, are incorrectly identified, or are placed in the wrong locations.

Although each of these three worlds presents slightly different problems, the same basic data fusion approach is applicable in all three areas. Therefore, this framework document addresses all three of these areas.

Definition of Data Fusion

In the defense area, the term "data fusion" has both a broad and a narrow definition. The narrow definition of data fusion is "the continuous process of assembling a model of the domain of interest from disparate data sources."

However, others use the term in a broader sense to cover a three layer hierarchical set of functions:

- First Level. At the first (lowest) level there is target identification and positioning, which depends on both the sensor and target characteristics. (The narrow definition only covers this level.)

- Second Level. The second level is situation analysis, which tries to infer what is happening by integrating the actions of multiple targets.

- Third Level. The third level is threat analysis, i.e., determining the seriousness of the developing situation and how to respond to it.

This framework will take the broader perspective and address all three levels. However, it will not address the image processing and signal analysis algorithms involving the individual sensor inputs.

Summary

Data fusion involves the integration of data from multiple sensors to provide a better analysis and evaluation of a situation that would be possible using data from only one sensor. "Better" in this case can have three components — more precise, more reliable, and faster.

- More precise may mean that you can determine that the vehicle is a tank transporter carrying a T-72, rather than just a loaded tank transporter.

- More reliable results reduce the uncertainty. For example, you may be have a .95 confidence level that the target is one of two types of vehicles versus a .7 confidence level or .8 confidence that it is one of four types of vehicle.

- Finally, a faster analysis of the situation is important if your required response time is very short and/or if you are dealing with multiple targets.

Data Fusion Approach

There are several different approaches to data fusion, including statistical classification, feature analysis, or neural nets:

- Statistical classification, the first approach, involves doing a single comprehensive analysis of the target data and determining the type of target and/or behavior.

- Feature analysis, the second approach, involves finding matches on specific features of the target. These features can be almost anything, such as a gun barrel, tracks, a turret, height, etc.

- Neural nets, the third approach, involves using neutral networks to find patterns.

All of these approaches involve some form of pattern (or template) matching. Traditional work with sensors just matches a pattern for a single sensor type and position. Data fusion involves matching multiple patterns from multiple sensors of either different types or at different locations.

The pattern matching analysis, i.e., the identification of the target type, may involve Bayesian, Dempster-Shafer, or voting algorithms.

A number of studies have suggested that the best results are obtained using data fusion at the feature level, as opposed to either the statistical classification or the neural nets level. Furthermore, the results are better if the features are independent of each other, rather than related.

Problem/Approach (high-level)

In general, data fusion represents a broad class of problems, i.e., information integration problems. Often this data comes in through sensors; but once it has been entered, its source is irrelevant. This type of problem is present in many domains, including RSTAKA, NPT (Non-Proliferation Treaty) verification, conventional arms control, health care (such as patient monitoring), manufacturing process control and adaptive control, and environmental monitoring.

Without further analysis, these domains are simply independent, possibly related areas. To select which ones to emphasize, we need a set of underlying dimensions along which they can be clustered and prioritized. The following is a set of dimensions:

- Real time control vs. monitoring.
- Database characteristics.
- Degree of uncertainty.
- Characteristics of the target markets.
- Portfolio analysis.

Real time control vs. monitoring

Real time data fusion is more complicated because it is operating under a time constraint. This can limit the amount of processing that can be done or require a more powerful system. Whether or not time is critical usually depends on whether the system is actually trying to control its environment or simply monitoring what is happening for later reporting. Real time data fusion is necessary if you are trying to control a battlefield or a manufacturing process. On the other hand, real time response is not necessary if you are only collecting data for later analysis, e.g., monitoring environmental pollution or traffic flow. In some cases, you may store the raw data and do both the data fusion and additional analysis later. In other cases, you may do the data collection and fusion initially and store only the processed data for later analysis.

Database characteristics

There are two key database characteristics (size and complexity) that vary by application area. However, there is some overlap, so size and complexity do not uniquely define an application area.

The size of the database is driven by the number of measurements being taken, which depends on the number of sensors and the sampling rate. The number of sensors depends on the area of coverage and the accuracy needed. The sampling rate may be determined by the required accuracy, the rate at which the environment or the measured characteristic is changing, and the predictability of the changes. The size is also affected by the type of measurement, e.g., a single value such as a temperature or pressure, or an image.

Complexity relates to the number of object types and the relationships among them (i.e., the number of fact types) in the database.

Size and complexity are essentially independent dimensions. Size drives the system's storage requirements, while complexity drives the processing power required by the system.

Degree of uncertainty

The degree of uncertainty is important because the basic purpose of data fusion is to reduce the degree of uncertainty about the environment, and in many cases, to determine how to respond to it. This uncertainty may involve the actual measurement(s), the situation that can be inferred from the measurement(s), how the situation is changing, or how to respond to the situation. Data fusion can improve the first three types of uncertainty, which may help in deciding how to respond if the situation involves some type of real time control.

Using multiple sensors, it may be possible to infer that there is a problem with one sensor if the sensor is providing readings that are inconsistent with the other sensors. More information from different types of sensors can reduce uncertainty about the situation because one type of sensor can often compensate for weakness in another type of sensor. However, this requires the ability to infer the meaning of one signature, given other signatures from other types of sensors, which is a more complex processing and analysis problem than simply analyzing and inferring from a single signature — e.g., developing joint templates from different types of signatures for each target type.

Characteristics of the target markets

Several characteristics affect the desirability of a market or application area for data fusion:

First, the size and growth rate of the market (i.e., need) are important. However, the resources and level of funding for the market are also important. A smaller, slower growing market may be more desirable, if it has more funding.

Second, the receptivity of the market to new data fusion technology is also important. Other things being equal, a market or application area (i.e., customers) that is more receptive to the required technology will be a much easier market than one that is resisting the technology.

Finally, the level of DOE and DOD funding is also important. DOE interests determine the basic Sandia missions and the extent to which they include data fusion and sensor activity. Also both DOE and DOD interests are a significant determinant of funding levels.

Portfolio analysis

Portfolio analysis is important because it is desirable to cluster data fusion work in areas that leverage each other. This is why understanding the similarity of the different data fusion areas is important. Focusing too narrowly on a specific area risks missing potential synergy across areas.

Methodology (high-level)

The methodology used to support this data fusion work involves a natural language oriented information modeling. As a user describes and explains (orally or in writing) a problem or a set of requirements, the natural language statements are mapped into a formal information model.

Integration of requirements and applications is done by combining these formal information models. During the development process, an information model is validated with the users, checked for consistency, and algorithmically grouped into a set of Fifth Normal Form tables for subsequent database design and implementation. Since the emphasis of this work is on the information integration aspects of data fusion, this section concentrates on developing the information model, not on the subsequent implementation steps for converting it into a database.

Information modeling allows a precise description of the type of information that is needed. This information may involve target and feature characteristics, sensor or signature characteristics, templates or patterns to be used to identify various types of targets, and actual signatures obtained by specific sensors at various locations. Major parts of the information models and much of the actual data can be effectively reused. In some cases this reuse will benefit a specific application area. In other cases, there is much broader synergy that benefits a number of areas.

An information model consists of a set of facts, examples, and constraints. A fact type consists of two object types (usually different types of objects, but they could both be the same type of object) and the role or relationship they have with each other. For example, an employee works in a department or a sensor is located at a position. For each fact there is an inverse fact, e.g., a department employs an employee or a position is the location of a sensor. This allows every relationship between two object types to be read in either direction and to convey meaningful information.

Information Modeling Methodology

This section provides an introduction to the concepts used in natural language-based information modeling. The information model can be represented in either of two ways — verbally or graphically.

- The **verbal representation** can be read, critiqued, and corrected by anyone who knows the subject matter, with virtually no explanation of the methodology.

- The **graphical representation** shows the relationships among the entity types more clearly and concisely, but it does require a few minutes of explanation to be understood. After reading this section, a person should be able to read and understand, although not construct, most of a graphical representation of an information model.

Concepts Covered in this Section:

- sentence
- elementary sentence
- fact
    - fact type
    - fact instance
- entity
    - entity type

entity instance

- label type
- role/verb
- constraints

  total

  uniqueness

Verbal Representation of Information Model

A **sentence** is simply a natural language statement by a user describing some aspect of the problem area. It may be simple, describing a specific example or instance — "Part X weighs 10 pounds," — or complex — "Part X, which was designed by John Smith last year, now sells for $100 and comes in red, green, and blue."

Any complex sentence can be decomposed into **elementary sentences**: "Part X was designed by John Smith." "Part X was designed in 1994." "Part X in 1995 sells for $100." Note that in this case, an elementary sentence is not binary: "Part X in 1995" does not provide the price. "Part X sells for $100" does not specify when, but the price may change over time. The other binary alternative — "1995 sells for $100" — has even more problems. The above sentences were all examples of specific instances, but they could equally well have been done in terms of types: "A part is designed by an engineer." "A part sells for a price in a year." "A part has a color."

The initial problem statement from the user is often a narrative consisting of simple and complex sentences referring to both types and instances. The information modeling methodology provides a way to decompose the problem statement into elementary sentences and formally model them to unambiguously identify all of the relationships and constraints in a way that the user can review them to verify or correct them.

The user describes the problem in natural language sentences: some are already elementary sentences while others are complex sentences. The complex sentences are decomposed into the corresponding elementary sentences. Many elementary sentences are binary, but they do not have to be. The key criterion is that an elementary sentence cannot be decomposed into more basic sentences without losing information, as shown in the previous part, date, and price example.

A structured sentence, sometimes called a **fact**, has a very specific form. It consists of two **entity types** (such as person, part, or department) that are related by a **role**, usually a **verb phrase** (such as designs, works in, or is responsible for). Examples of fact types include the following:

- A person designs a part.
- A person works in a department.
- A department employs a person (the inverse of the previous fact).

For each fact type there can be many **fact instances**, such as "Bill designed part 1234," or "Sam works in Engineering."

To completely capture all of the required information, a deep structure sentence or fact has a specific form. It specifies the first entity type, its identifier or **label type**, several examples or **instances** of that label type, a verb phrase, and another entity's set of information (i.e., entity type, label type, and label instance). Although label instances are sometimes called examples (in the sense that they are examples of entity types), the information modeling methodology really requires examples of facts or fact instances.

Entity type:

Person

Label type:

SSN

Label instance:

123-45-6789

Verb:

works in

Entity type:

Department

Label type:

Department name

Label instance:

Engineering

Fact instances or examples are critical because they explicitly define the data constraints, which the DBMS must enforce. Let's explain the **constraint types** using specific examples for the fact pair: "A person designs a part" and "A part is designed by a person."

The **total constraint** tells whether or not every entity instance of a specific type must participate in the fact type. Must every person design a part? No, so the first fact/role is not total. However, must every part be designed by a person? If we assume the answer is yes, then this fact/role is total. If we know anything about a part instance, we must know the person who designed it. (Database experts will recognize that this is a mandatory attribute for an entity; but the user has remained insulated from that designer view.)

In an actual modeling session, someone may raise the issue that we buy some parts from suppliers and for those parts the designer is unknown and probably irrelevant. In other words, for some parts, one set of facts apply, while for other parts, a different set of facts may apply, although all parts share a common set of facts. This distinction defines the subtype-supertype relationship. The supertype (part) has a set of facts that are common to all of the subtypes (designed part and purchased part). The subtypes are distinguished from each other by the unique set of facts that apply to each subtype. All of the common facts (except the identifier) are removed from the subtypes and are associated with the supertype.

To determine another important constraint — the **uniqueness constraint** — requires an additional example. Consider the following examples for the fact "a person works in a department."

	Person	Department
1.	Sam	Eng
2.	Mary	Mfg
3.	Bill	Eng
4.	Sam	Finance
5.	Joe	—
6.	—	Accounting

When shown the previous six examples, the user can quickly determine which ones are good:

- Examples 1 and 2 are good because there is no overlap; they are two independent fact instances.

- Example 3 is good because a department (Eng) can have more than one person in it.

- Example 4 is incorrect, however, because a person (Sam) can only be in one department. This defines a uniqueness constraint, an specific instance of a person can only appear once in this fact type.

- Examples 5 and 6 simply document the totality constraints described above. Example 5 is incorrect because every person must be in a department, i.e., the total constraint. Example 6 simply verifies that departments may be created and other data collected about them before people are actually assigned to them. However, this is only a business rule constraint, not a physical constraint, so another company could decide that they wanted to consider example 6 as incorrect.

After analyzing these examples, a more precise statement of the facts is possible. These were the initial facts:

- A person works in a department.
- A department employs a person.

Considering the examples, the more precise facts are as follows:

- Every person must work in one department.
- A department may employ one or more people. (Note: The zero people case is implied by the "may" in this example.)

The possible uniqueness constraints are that the object on the left may be unique, the object on the right may be unique, each object may be unique, or the combination may be unique. An example of fact with the combination is "a person is assigned to a project." A person can be assigned to many projects and a project can have many people, but you would not assign Sam to project X twice.

Graphical Representation of Information Model

For completeness, the rest of this section briefly describes the graphical representation of the information model. In Figure 1, the solid circle represents an object or concept in the real world, such as a person, a department, a part, or a release status. Dashed or dotted circles represent data objects that identify or further describe real objects, such as employee name, social security number, or release code. Boxes or rectangles represent the roles played by one object type with respect to another. The two boxes together indicate that two roles are complementary — a person works in a department and a department employs a person. With appropriate naming, facts in graphic model can be read as sentences.

Figure 1 shows several basic facts (in both directions) and their constraints. The facts shown include: "a person is identified by a SSN," "a person works in a department," "a person designs a part," "a person is assigned to a project," and "a part has a current release status." (Note: The model must specify "current" release status because a part will have many release statuses over time.) The constraints are also shown. The V indicates a total constraint and the line over a role indicates uniqueness. Obviously there are additional constraint types and symbols, but this should provide the reader with a general understanding of the graphic model representation. The neutral data model that can be generated from the information model (in either its verbal or its graphical representation) can be represented in any of the traditional data modeling notations.

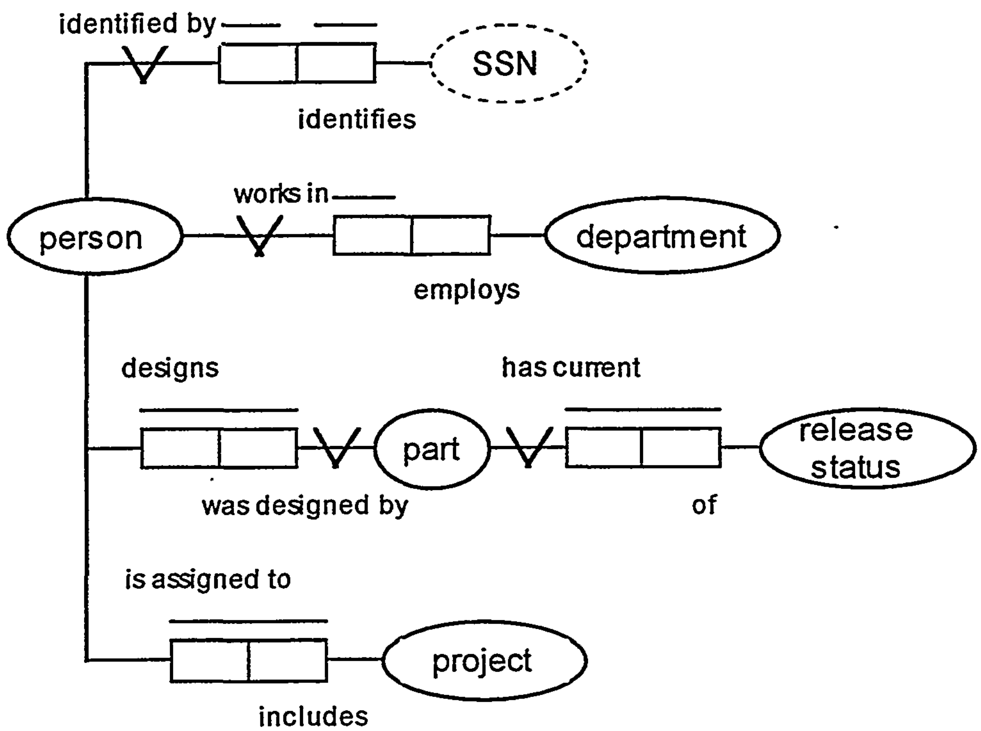


Figure 1. Example of graphic representation of information model

Data Fusion Information Models

Traditionally, data fusion models, when they were explicitly done, focused on application-specific types of objects, such as tanks, planes, robot arms, workpieces, and machine tools. At one level these models are good because they focus on the specific types of objects and relationships a set of users knows and understands. However, by only focusing at this specific level, we are losing some of the synergy and benefits that can occur by looking across application domains. To capture some of this synergy, the next section suggests a high-level, generic information model. Information models in the following parts apply this generic model to specific application areas.

The basic generic information model has five object types — targets, behaviors, environments, signatures, and sensors. The basic facts are:

- A target may be located in one or more environments.
- A target includes one or more features.
- A target may exhibit one or more behaviors.
- A target exhibiting a behavior in an environment generates a signature.
- A feature exhibiting a behavior in an environment generates a signature.
- A signature can be detected by one or more sensors.
- A sensor can detect one or more signatures.

These basic facts are equally applicable for most data fusion application domains such as defense, manufacturing, health care, or environmental monitoring.

In the defense domain, a target such as a tank or a missile carrier may be located in desert, swampy, or rolling terrain. The environment may also be dry, raining, snowing, or foggy. Certain types of targets may not be able to operate in some environments. Targets can exhibit behaviors such as moving, rotating a turret, firing, erecting a missile, or launching a missile. Again certain behaviors are only appropriate with certain types of targets. A target exhibiting a behavior in an environment generates a signature. A launcher firing a missile generates a number of signatures — visual, radar, thermal, acoustic, and seismic. However, the appearance of these signatures may be affected by the environment in which they occur. For example, the seismic signature will be different if the launching occurs on hard ground or softer marshy ground, although there is a limit on how soft the ground can be and still allow the launcher to move in it.

Parts of specific information models described in other parts of this report (and in other reports from this LDRD project) can be directly related back to this core generic information model, i.e., this key set of facts. This connection provides the synergy across the various data fusion areas.

Conclusion

This section has provided a brief introduction and framework for data fusion and the information modeling approach that was used by this project.

- First, we explained why data fusion is important and the use of single and multiple sensors, types of sensors, and levels of intelligence being built into sensors.

- Second, we reviewed types of data fusion applications in terms of environments (designed, real, and hostile worlds) and their characteristics, and suggested that there is significant overlap in the application characteristics, which could provide opportunities for synergy across application types.

- Third, we briefly described the information modeling methodology used and identified a basic generic model with five common object types — targets, behaviors, environments, signatures, and sensors.

The remaining sections discuss data fusion within a specific context and describe some of the related models.

Intentionally Left Blank

2. Data Fusion Lessons for an ATR/SAR Project

Introduction

This section of the report shows the potential benefits of the information modeling approach to data fusion. It shows these benefits within the context of a previous, successful ATR/SAR (Automatic Target Recognition/Synthetic Aperture Radar) project. It does this in several ways:

- First, it shows the relationship between the specific information model developed bottom up for the specific project and the more generic, top down model developed under the data fusion LDRD.

- Second, it shows how two areas included in the generic model, but which were outside of the scope of the ATR/SAR project, can be easily integrated into the specific model as extensions. These two areas are target behavior and multiple sensors.

- Finally, it describes how the data fusion model can provide a reusable basis for other projects in the future.

Why revisit a successful project such as ATR/SAR? You can always learn something from and improve an unsuccessful project. A more significant test is to determine whether the proposed approach or changes improve an already successful project, such as ATR/SAR. Given the generic data fusion information model developed under the LDRD, the ATR/SAR information model could have been developed much quicker. Furthermore, it would have been even more reusable, although the current version of the ATR/SAR information model is still general enough to be very reusable.

The ATR/SAR model was one of several low-level bottom up information models developed for different types of applications that were used as the basis for developing the generic data fusion information model. However, the approach of customizing a generic or high-level model is also a very productive way to develop information systems, and a generic data fusion model is an ideal starting point for many other data fusion and sensor projects.

This section reviews the results of the ATR/SAR project, describes the possible extensions from the data fusion model, and concludes with lessons for other projects.

ATR/SAR Project

The overall purpose of the ATR/SAR project was to develop templates for recognizing various types of targets in various orientations on different terrains. The database and its related applications support the data collection and analysis functions. Since the purpose of the project was to develop templates for a specific type of sensor, the targets were of known types, positions, and orientations and were stationary.

In the past, the ATR/SAR group had managed its image data using a more generic image management system from a commercial vendor. However, there were problems with both responsiveness and functionality, even with the lower past data volume. A new generation of hardware technology allowed the data collection to increase by over two orders of magnitude. This required a much more effective and focused information system, for which the information modeling work was done. Keeping track of, retrieving, and analyzing this large volume of data was critical to the project's success.

The five types of data to be managed included the following:

1. Site/Survey management involved identification and description of each site. This included the site survey data for permanent terrain features, temporary characteristics such as types of vehicles and their positions, orientations, configurations, and calibration markers. The same site could be used many times with different sets of target vehicles and calibration markers.

2. Flight management involved a sensor-equipped aircraft flying a specified flight path around the site collecting many images of the area. By knowing the flight path and the time an image was taken, and by finding the calibration marker on the image, the system can calculate a "normalized" view of the site so that it can map from image coordinates to site coordinates (where there should be a target) for additional analysis.

3. Image management identified and kept track of the characteristics of each image. Data of interest may include (when the image was made) the equipment that was used and the targets (type, position, and orientation) in the image. Since the program may generate 40,000 images, this type of management and retrieval is important.

4. Patch management was used because for more detailed analysis, small patches, which include a target and its immediate area, need to be isolated. With perhaps ten targets per image, there may be 400,000 patches to manage as input to the analysis routines. Each patch needed to be identified, stored, and characterized with respect to the image from which it came, where on the image, and the target (type, position, orientation, and configuration) it includes.

5. Test/analysis/template management was the ultimate focus of the project. The purpose of the analysis was to determine a template for finding a type of target with a specific orientation. For a specific analysis, certain patches were used as inputs, a template was calculated, and other patches were used to validate the template. All of this information needed to be stored and tracked for each analysis. Later in a target recognition context, this template would be used to analyze an image of an unknown area to find and identify specific targets.

For the Sandia National Laboratories' Information Technology group, the scope of the work included developing the information model and designing and implementing the database. The actual image analysis applications and template development were outside of this scope and were done by a specialized image analysis group. However, management of the images and patches to be analyzed and the templates that were developed were within its scope.

The technical approach used to define the requirements and design the database was the natural language-based information modeling methodology (described in the previous section), which develops a fact-based model of the information requirements. The various types of facts are identified by interviewing users, analyzing reports and screens, and to an increasing extent reusing existing facts. The Appendix for this section provides a summary version (10 to 15 percent of the full model) of key parts of the ATR/SAR Information Model.

Reusing existing facts reduces the cost and time for developing a new system and helps ensure that related systems (i.e., those that share a common set of facts) can be more easily integrated or interfaced with each other. For example, much of the work for this SAR project could be used to manage images and develop templates for other types of sensors.

Discussion

As a context for this discussion, recall that the five common or generic types of objects for most data fusion applications are target, behavior, environment, signature, and sensor.

The purpose of ATR/SAR was to develop templates (actually algorithms to develop templates) to identify certain types of objects or targets, such as various types of vehicles. The basic information structure involved a site where a lot of imaging work was done, and scenes, which were parts of a site. Objects were placed throughout the scene and images were made of the scene. The objects included both extended targets (e.g., vehicles) and point targets (e.g., calibration devices, which permitted accurate mapping between points on the ground and positions on the image).

The scene was further divided into groundcells, which had both static characteristics (e.g., rocky, sandy, and contour — strike and dip angles) and dynamic characteristics (e.g., water, snow, vegetation, and camouflage layers). Some groundcells were occupied by objects. If the object was an extended target such as a truck or a tank, the target type had a reference point, a length, width, and height. Each instance of a target had a type, a position, an orientation, and a configuration. Target configuration included factors such as whether doors or hatches were open or closed and whether a vehicle was loaded or empty. All of this data was collected by surveying the scene before the imaging flight. During the flight, the scene was imaged from many perspectives for later analysis.

While the above model and the supporting applications were developed for building SAR templates, virtually the entire model is appropriate for developing templates for any static images. The site-scene-groundcell structure is directly applicable, as are the terrain data (which corresponds to environment in the general form of the model) and the target data, including type, position, orientation, and configuration. The tophats and corner reflectors were specific types of devices for SAR, so they would need to be generalized to support other types of sensors, but this would be a relatively minor enhancement.

Similarly, the image (or signature in the more general form) would need to be extended, initially with only a single fact to specify the type of signature. This was not needed in the current model because only SAR images were being collected. This extension would allow the system to keep track of different types of images (or signatures) such as SAR or thermal or multispectral images.

Behavior was the only one of the five common elements that was not included in ATR/SAR or could not be easily added. To accurately and effectively include dynamic behavior in the model would require a significant amount of work. However, this work would primarily, if not exclusively, add new objects and facts to the information model. It does not appear that it would significantly change or restructure the existing model. This means that such extensions would have minimal impact on existing applications that were using the current information model. To be fair, it should be noted that dynamic behavior was completely outside of the scope of the original ATR/SAR project for which the information model was developed.

Finally, extending the ATR/SAR work to support multisensor data fusion would require changing the algorithm — possibly only extending it. With data fusion, instead of developing a single template for a given target (with a position, orientation, and configuration) in a specific environment, there would be a template for each type of signature and an algorithm for combining the results. However, the same algorithm might be appropriate for each signature individually.

Summary

A generic data fusion model would have helped the ATR/SAR project, although the project was very successful without such a starting point. Conversely, given the generic data fusion model, it is easy to see how the ATR/SAR information model could be extended to cover the more general data fusion case. It already explicitly includes the model components for target and environment. It implicitly includes signature and sensor since it only dealt with SAR, but a simple extension to the model would make these components explicit. Behavior is the only part of the generic model that is missing and it was outside of the scope of the original project.

Appendix for Section 2 - Summarized ATR/SAR Information Model

This appendix provides a summarized version of the ATR/SAR information model, concentrating on the entities and facts needed for the discussion in this section of the report. The complete model included more than 450 facts.

Entity Types Included in this Appendix:

- site
- scene
- survey-reference-point
- scene-center
- groundcell
- Dynamic-groundcell (DGC)
- Occupied-dynamic-groundcell (ODGC)
- camouflage-layer
- object-type (equipment)
- point-target
- object-instance

From the data fusion study, the following entity types could be added to or, in some cases, replace some of the above entities for a more general information model:

- target
- behavior type
- behavior instance
- environment
- sensor type
- sensor instance
- signature

Definitions of the Entity Types:

Site

A site (which is a general area such as an "abandoned airfield") must be identified by a site-id (MW).

A site may include one or more scenes.

Scene

A scene must be identified by a scene-id
(scene-id = site-id + scene-sequence-number).

A scene must have one scene-center.

A scene must have one survey reference point (SRP).

A scene must be imaged on one or more flights.

A scene must include one or more groundcells.

A scene may be captured on one or more images.

Survey-reference-point (SRP)

An SRP must be identified by one SRP-id.

An SRP must have one latitude angle.

An SRP must have one latitude direction.

An SRP must have one longitude angle.

An SRP must have one longitude direction.

An SRP must have one elevation.

Scene-center

A scene-center must be identified by one scene-id.

A scene-center must be for one or more scenes.

A scene-center must have one latitude angle.

A scene-center must have one latitude direction.

A scene-center must have one longitude angle.

A scene-center must have one longitude direction.

A scene-center must have one elevation.

Groundcell

A groundcell must be identified by one groundcell-id
(groundcell-id = scene-id + groundcell-sequence-number).

A groundcell must occur within one scene.

A groundcell may have one ground-type or scene-background
(concrete, grass, sand, water, etc.).

A groundcell may have one or more comments.

A groundcell must have one coordinate
(x, y, z) or (distance, angle, angle).

A groundcell must have one offset distance (a linear-dimension).

A groundcell must have one offset angle (an angle in degrees).

A groundcell may have one strike angle.

A groundcell may have one dip angle.

Dynamic-groundcell (DGC)

A DGC must be identified by one DGC-id
(DGC-id = groundcell-id + time).

A DGC may have a water-level
(dry, rain, dew, etc.).

Occupied-dynamic-groundcell (ODGC)

An ODGC must be identified by one ODGC-id
(ODGC-id = DGC-id + object-instance-id).

An ODGC must include one object-instance.

An ODGC must have one pitch angle.

An ODGC must have one roll angle.

An ODGC must have one yaw angle.

An ODGC may include one or more camouflage-layer.

An ODGC must include one configuration or object-state
(structured like camouflage-layer, below)
(pre-defined set of values).

An ODGC may include one or more deployment-types
(none, top, side, top and side).

Camouflage-layer

A camouflage-layer must include one camouflage-layer-number.

A camouflage-layer must include one camouflage-layer-type
(pre-defined set of values).

A camouflage-layer may be for one or more ODGCs.

Object-type

An object must be identified by one object-type-id
(REF, PTT, NRA, VEH, EXP, REF/TOPHAT).

An object-type must be one of two subtypes
(point-target or extended-target).

An object-type must be described by one object-description.

An object-type may have one object-reference-point (ORP).

An object-type may be the type of one or more object-instances.

An object-type may be located in one or more ODGCs.

An object-type may have one or more moveable-parts
(related to configuration).

An object-type may have one or more configurations.

Point-target

A point-target must be identified by one object-type-id
(a subtype identified by the supertype id).

A point-target must be of type one point-target-type
(tophat or corner reflector).

A point-target may have one radar-cross-section (RCS) model.

Object-instance

An object-instance must be identified by one object-instance-id
(object-instance-id = object-type-id + object-sequence-number)
(object-instance-id = object-type-id + serial-number).

An object-instance must be of one object-type.

3. Peace/Truce Management Framework

Peace/Truce Management

Introduction

In 1992 when this LDRD was proposed, John Taylor (5335) suggested truce management as a possible area to consider for this type of data fusion modeling. Although at the time we could not find an immediate customer, the area of truce management did seem to have growth potential and could be built on two Sandia competencies — information and sensors. Therefore, this part of the report is a preliminary exploration of truce management and identifies some ways that data fusion modeling could support this area. Without an immediate customer to provide detailed requirements and subject matter expertise, the information model at the end of this section is only suggestive of the direction to go and is not intended to be complete.

What is Peace/Truce Management?

Since the breakup of the Soviet Union, U.S. defense requirements have changed. There is less emphasis on major strategic systems and more importance must be attached to conventional arms, instability, and major regional conflicts (MRCs) in many parts of the world.

These conflicts and problems, normally involving only conventional arms, may range from relatively unsophisticated guerrilla warfare to traditional, open, declared warfare with very sophisticated conventional weapons to civil wars where one or both sides may or may not have sophisticated weapons.

Even the nature of low-intensity conflict is changing. Traditionally, guerrilla warfare involved an insurgent group (with or without external support) fighting a central government; but in the case of a failed state such as Somalia (and in the future, perhaps parts of the Former Soviet Union), there may be little, or a noneffective, central government. In some respects, this situation is similar to warfare among rival gangs, which may require different forces, tactics, and much greater intelligence. In any of these cases, a new concept called Peace Management or Truce Management can play a major role in stabilizing or terminating a hostile situation and providing time for diplomacy to work out a solution.

There are many situations where this concept of peace/truce management could be applied in the Middle East, the Balkans, Africa, and perhaps even in some parts of the former Soviet Union. However, for this approach to work, both sides (or all sides if there are more than two) must want to stop the conflict, with distrust and insecurity being the main barriers. If for any reason one or more sides wants to continue the fighting, this type of truce management may not work or will be much less effective. If an international "peacekeeping" effort were attempted in this case, some of the same technology could support this more active type of intervention but it would be something different than "truce management," perhaps "peacemaking" rather than "peacekeeping."

Peace/truce management involves the extensive use of sensors to monitor what is happening in a truce or cease fire area and detect any anomalous behavior. For example, if the terms of a truce called for the withdrawal of tanks and heavy equipment by both sides, sensors could be placed throughout the truce area to detect the presence of tanks or other unauthorized equipment. Who does this monitoring (the parties to the dispute, an acceptable third party, or an international organization) and how the data are collected and used are distinct political issues that need to be addressed for any specific situation.

However, the focus of this section is on the information model and the type of data needed to support a peace/truce management function. In other words, the section addresses the underlying enabling information integration technology, rather than the detailed implementation in a specific political setting. Also the more technical issues about sensor technology are beyond the scope of this paper.

Examples of some technical issues that can have an impact on the political arrangements for truce management include the following:

1. The power supply and unattended life of the sensor and whether or not the sensor can be turned off to conserve power.

2. Communications requirements from and to the sensor.

3. Remote interrogation and command capability of the sensor.

4. Authentication.

5. Tamper resistance of the sensor, which may be determined by inspection or remote interrogation.

6. The capability of the sensor to operate in a hostile environment, e.g., rain, snow, dry, heat, and cold.

There are two distinct types of applications for data fusion and analysis systems in this peace/truce management context:

- The first application involves the type and placement of sensors. Given the terrain, the types of equipment and movement to be monitored, the characteristics of that equipment, and the characteristics of the sensors, it is possible to determine an effective distribution of sensors, i.e., how many of what types of sensors should be placed and where they should be placed to maximize the probability of detecting increased activity that might be associated with truce violations on a militarily significant scale.

Not only must an effective distribution of sensors be determined, but the involved parties must be convinced that it will be effective. This could involve extensive analysis, simulation, and testing of various alternatives, i.e., the type of work being done by the Cooperative Monitoring Center at Sandia National Laboratories.

- The second application is the more traditional data fusion problem of how to analyze and integrate the data from all of the sensors to determine what is happening and if any increased activity might contribute to a violation.

  When to consider and label something as a violation is essentially a political rather than a technical decision. The technology can only tell you what is happening, e.g., several sensors have picked up tank signatures. The technology cannot tell you how to interpret it, e.g., (1) the tanks are withdrawing as agreed to, or (2) they should not be there so there is a violation, or (3) they should not be there but to facilitate negotiations to get them out, their presence will not explicitly be called a violation.

  Ideally, this analysis should be done in real time so that there is time to take corrective action. "Corrective action" may involve deploying additional sensors or on-the-ground inspectors to verify what is happening, notifying the parties to the agreement of the situation, or at the more active extreme (such as "peacemaking"), carrying out some enforcement action.

Scope of the Problem

Range of Alternatives for Truce Management

In its broadest scope, the problem is very large and complex. It requires considering many different types of equipment and behaviors or characteristics of that equipment, many different types of sensors and how they would see those characteristics and behaviors, and how those sensor perceptions would vary in a wide range of environments (e.g., terrains and weather conditions). The development of an underlying information model to truce management, while difficult, may be much easier than collecting the actual data. However, the information model provides a framework and the requirements to guide the data collection and storage process.

Examples of the range of alternatives indicate the complexity of the problem. The sensors may be imaging or nonimaging, use ground-based, airborne, or satellite platforms, be used continuously or randomly, and be in positions known or unknown to the parties. If their positions are to be unknown, how do you convince the parties of the effectiveness of the deployment and how do you install and maintain the sensors without exposing their positions? What behaviors (and corresponding signatures) are clearly acceptable, clearly unacceptable, or ambiguous enough to require additional information/confirmation by on-the-ground inspectors? What are the procedures for these inspections?

In some respects, negotiating a truce becomes more complicated because it must include the necessary terms and conditions to allow for the effective deployment and use of the sensors and inspection teams.

University of New Mexico Information Model

A very basic expert system prototype based on an early information model was developed for the University of New Mexico seminar on decision support and expert systems. It considered only traffic anomalies with respect to traffic volume, load, and direction, and reported conflict to prioritize geographic areas of concern. For example, these areas and the types of problems could help determine what types of "corrective actions" were needed and where.

The information model did not consider the additional real world complexity such as the effects of weather, attempts at deception and concealment, the fact that a sensor may see different types of terrain looking in different directions, and probability or confidence levels.

However, it was still useful for three reasons:

- First, it provided an initial, although very simple, structure that can be expanded and enhanced for a more complete prototype.
- Second, it allowed us to use a fact-based information model (using NIAM) and create the fact instances needed to develop a rule-based expert system prototype. So far these extensions have been done manually. But we now understand what is needed, so an automated tool could be developed or an existing tool could be extended to support this capability, which is needed for expert systems work. Although these extensions have been discussed, actual tool development or modification was not undertaken as part of this LDRD.
- Third, it provided a prototyping experience with an expert system shell (EXSYS).

The natural language-based information modeling methodology (described in Section 1) was used to develop the Truce Management information model included at the end of this section. The expert system prototype used only a small subset of these facts for its rule base.

Truce Management Scenario

A truce management system could operate in either of two ways or in some combined approach. At the more detailed level the system could take sensor input, probably from multiple sensors, and determine the specific type of target (tank, truck, or car) and its behavior. This could be done based on the input from one sensor or using data fusion multiple sensors (of the same or different types). In general, this approach is similar to a target acquisition and identification function. However, at a more general level, the system could only focus on analyzing and characterizing the overall situation and helping decide how to respond to it.

As an example of the more detailed approach, one could use a combination of radar, IR, acoustic, and seismic sensors to distinguish among a tank truck, a missile carrier, a tank transporter, and a tank. However, this would require very precise data and a lot of processing.

At the more general level, the issue might be simply to identify a heavy volume of traffic toward the border or truce line, especially if there were attempts to conceal it, and the pattern continued over several days. For this analysis it might not matter what types of vehicles were involved, only that the volume was much heavier than expected and it was moving in a critical direction.

A significant difference between the detailed and the more general approaches is that in the latter case, you are not targeting the vehicles for destruction, only for further investigation, so less initial precision is required. This means that the consequences of uncertainty are less severe. Also it implies that time and resources are available to gather additional clarifying information if necessary.

Another difference between the two approaches could involve fewer and/or less expensive and less capable sensors and/or less sophisticated processing of the sensor data. The complete set of alternatives can be specified by a 2×2 matrix with the two dimensions being sensor precision and result detail.

Results	Sensors	
	Low Precision	High Precision
General	1	2
Detailed	3	4

Case 1 involves low-precision (LP) sensors and only enough processing to provide general results, such as information on traffic patterns (e.g., volume and load) but not detailed data on individual vehicles. This approach is not only the least expensive, but also the least threatening to the involved parties. With low-precision sensors, the assumption is that the results could not be upgraded into actual targeting data, i.e., case 3 does not exist. The use of cheaper sensors would allow more sensors and therefore either coverage of a broader area or a more intensive coverage of the same area. Politically this approach might be the easiest to initially get accepted by the opposing parties.

Case 2 involves the use of high-precision (HP) sensors, but only reporting general-level results. This can occur in either of two ways — minimal processing that provides only general rather than detailed results or more complete processing to produce the detailed results but only reporting the summary or general results. The key characteristic of this approach is that it could easily evolve into case 4 simply by changing the processing or the reporting. It does not require changing or replacing the sensors placed throughout an area. In some respects, this is an advantage, but in other respects it could be considered a disadvantage by some parties. With high-precision sensors, unless the agreement was for case 4, the parties could not be sure how much information the monitoring group had — or if the data were shared, how much data their opponent had. Given this uncertainty, this approach might be politically difficult to get agreement on.

Case 3 — generating detailed results from low-precision sensors — is a null case. If the sensors provided enough precision and data for the detailed results, then by definition they would be considered high-precision and would become case 4. (Note: This is an assumption that may or may not be true. For example, using data fusion, could an array of low-precision sensors provide the type data needed to generate detailed results that could be used for targeting? If so, this complicates the sensor problem, unless the parties are willing to allow cases 2 and 4.)

Case 4 involves high-precision sensors and detailed results. This approach is the most expensive, although perhaps only marginally more so than case 2. However, it should provide the greatest precision, confidence, and reliability. A potential difficulty is that it may provide more than the parties are initially willing to reveal.

A hybrid approach could involve primarily case 1 — low-precision sensors and general results — and case 4 in a few particularly critical areas.

Information Model

This type of truce management may involve two types of monitoring:

- First, it may require monitoring the transportation network for the region to identify unusual traffic patterns, which could indicate prohibited activities such as heavy movements of equipment and/or personnel. This heavy movement could warn of a build up or confirm a withdrawal. It could also mean heavy refugee traffic, which could indicate prohibited expulsions or ethnic cleansing.

- The second type of monitoring would involve other types of key areas (i.e., nontransportation) where prohibited activities could occur, especially areas where such activities would have serious consequences for the truce agreement — for example areas where equipment could be stockpiled or training could be done for a future surprise offensive.

The rest of this paper concentrates on the transportation network and monitoring in this setting. Section 2, "Data Fusion Lessons for an ATR/SAR Project," described a more detailed information model, which is more applicable for placing sensors anywhere and monitoring any area for specific types of targets. This section focuses on the case 1 scenario, i.e., the general, less detailed sensors and results, because it is more likely to be accepted initially. Since there may be very little difference between some aspects of the detailed type 4 truce management and the target acquisition needed for enforcement actions, type 4 truce management may not be initially accepted by all parties to a conflict.

The rest of this section provides a brief narrative of the problem and then specifies a preliminary high-level information model, as a set of natural language sentences. For a more detailed description of the natural language-based information modeling methodology, see Section 1.

Narrative

The information model described next is a high-level preliminary model, which specific users could refine to focus on any one or combination of several areas. Major objects in the information model include the truce agreement, the area(s) covered and its borders, the transportation network, traffic patterns, targets, and sensors. Although the traffic network part of the model was developed for the road system, it was generalized to support other similar networks such as rail and water transportation. With some enhancements, it could also support air traffic, but the current work does not include this extension.

Facts about the truce agreement include the area(s) covered, the parties, the date it was signed, the date it became effective, and the types of actions that are permissible and/or prohibited. Some of these actions should include the placement and maintenance of sensors (probably by third parties) and permissible notification and inspection procedures. Facts about areas include boundaries, sites, transportation networks, and actions that can occur within the area. Information about the transportation network include its type (road, rail, or water), segments with starting and ending points, and capacities (volume and/or load) of a segment. Traffic patterns and sensors can be associated with any segment at any time. Targets generate the signatures seen by the various sensors.

This preliminary model can be extended with additional detail in several different directions. The model in Section 2 focused in greater detail on the target and sensor aspect. This model provides the major linkage points to that model, but does not duplicate the work done there. One of the benefits of the natural language-based information modeling approach is that it is very easy to link related fact-based information models.

Main Objects in the Truce Management Information Model

For the objects marked with a " * " there are corresponding objects in one or more of the following: the On-Site Inspection Model, the Manufacturing Model, or the ATR/SAR Model.

- Action-type
- Action-instance
- Actor
- Agreement* (Corresponds to treaty in the On-Site Inspection Model)
- Anomaly
- Area
- Boundary
- Clause*
- Country*

- Facility*
- Global-position*
- Measure-type*
- Measure-instance*
- Point
- Segment-instance
- Sensor-instance*
- Signature-type
- Signature-instance
- Signature-perceived
- Target-type*
- Target-instance*
- Traffic-pattern-type
- Traffic-pattern-instance
- Transportation-network-type
- Transportation-network-instance
- Vehicle-type* (Corresponds to object-type in ATR/SAR)
- Vehicle-instance* (Corresponds to object-instance in ATR/SAR)
- Vehicle-behavior

Possible additional objects from the On-Site Inspection model:

- Escort
- Inspector

- Inspection
- Inspectable-area
- Itinerary
- Mandate
- Path-segment

Definitions of the Main Objects in the Information Model

Action-type

Every action type must be identified by an action type code.

Every action type must have an action name.

Every action type must have a description.

Every action type must be defined by one or more action characteristics.

An action type may be the type of an action instance.

Action-instance

Every action instance must be identified by an action-instance-id.

Every action instance must be started at a date/time.

An action instance may have ended at a date/time.

Every action instance must have occurred within one or more areas.

An action instance may have occurred at one or more global positions.

Every action instance must involve one or more actors.

An action instance may have been caused by one or more actors.

An action instance may have victimized one or more actors.

An action instance may have caused one or more damages.

An action instance may have been directed at one or more targets
(targets may be persons or things).

An action instance may have had one or more consequences.

An action instance may have used one or more resources.

An action instance may have left one or more evidences.

An action instance may be related to one or more anomalies.

Actor

Every actor must be identified by an actor-id.

Every actor must have a name.

An actor may make one or more agreements.

An actor may be affected by one or more agreements.

An actor may initiate one or more action-instances.

An actor may be the target of one or more action-instances.

An actor may control one or more areas.

The subtypes of actor include: country, alliance, international organization, or faction.

Agreement

Every agreement must be identified by an agreement-id.

Every agreement must define an area.

An agreement may be signed on a date.

An agreement may specify an effective date.

An agreement may specify a termination date.

An agreement may specify a review date.

Every agreement must consists of one or more clauses.

Every agreement must be made by one or more actors.

Every agreement must affect one or more actors.

An agreement may specify one or more permitted action-types.

An agreement may specify one or more prohibited action-types.

Anomaly

Every anomaly must be identified by an anomaly-id.

An anomaly may be related to an action.

An anomaly may be determined by one or more evidences.

An anomaly may involve one or more actors.

An anomaly may have been initially detected at a date/time.

An anomaly may have been detected by one or more persons.

An anomaly may have been declared by one or more actors.

Area

Every area must be identified by an area-id.

An area may be defined by one or more agreements.

Every area must be limited by one or more boundaries.

An area may include one or more sites.

An area may include one or more transportation networks.

An area may be the location of an action-instance.

Boundary

Every boundary must be identified by a boundary-id.

Every boundary must partly define one or more areas.

Every boundary must include a boundary description.

(Currently, a boundary is simply a text description.)

Clause

Every clause must be identified by a clause-id (agreement-id + clause-number).

Every clause must be included in one or more agreements.

A clause may involve one or more countries.

A clause may involve one or more areas.

A clause may specify one or more action types.

A clause may specify one or more inspection procedures.

Country

Every country must be identified by an actor-id.

Every country must have a country name.

A country is a subtype of actor.

Facility

Every facility must be identified by a facility-id.

Every facility must be of a facility type.

Every facility must be located in a country.

Every facility must have a global position.

A facility may be located in an area.

A facility may be managed by one or more actors.

Global-position

Every global position must specify a latitude.

Every global position must specify a longitude.

Every global position must specify an altitude.

Measurement-instance

A measurement-instance must be identified by a measurement instance id.

A measurement-instance is of a measurement type.

A measurement-instance must have a value.

A measurement-instance must be taken using a sensor-instance.

A measurement-instance must be taken at a date/time.

A measurement-instance may be made under one or more environmental conditions.

Point

Every point must be identified by a point-id.

Every point must be defined by a global position.

A point may be the location of one or more sensors.

A point may be the location of one or more action instances.

A point may be the start of one or more segment instances.

A point may be the end of one or more segment instances.

Segment-instance

Every segment-instance must be identified by a segment-instance-id.

Every segment-instance must be of type segment type (i.e., road, rail, or water).

Every segment-instance must be located within one or more transportation network instances.

Every segment-instance must start at a point.

Every segment-instance must end at a point.

A segment-instance may have a maximum volume capacity.

A segment-instance may have a maximum load capacity.

A segment-instance may be monitored by one or more sensor instances.

A segment-instance may be the location of one or more action instances.

Sensor-instance

Every sensor-instance must be identified by a sensor-id.

Every sensor-instance must be of type sensor-type.

Every sensor-instance must have a status.

Every sensor-instance must be placed at a position at a date and time.

A sensor-instance may be removed from a position at a date and time.

A sensor-instance may be used to make one or more measurement-instances.

Transportation-network-instance (TNI)

Every TNI must be identified by a TNI-id.

Every TNI must consist of one or more TNI-segments.

Vehicle-type

Every vehicle-type can exhibit one or more vehicle-behaviors-types.

Every vehicle-type must have a range.

Every vehicle-type can operate in one or more terrains.

Every vehicle-type has a maximum speed.

Every vehicle-type has a maximum load weight.

Vehicle-instance

Every vehicle-instance must be identified by a vehicle-id.

Every vehicle-instance must be of a vehicle-type.

A vehicle-instance may be located at a position, at a date and time.

Vehicle-behavior-type

Every vehicle-behavior-type must be identified by a VBT-id.

Every vehicle-behavior-type must be described by a VBT-description.

Every vehicle-behavior-type must result in one or more generated-signatures.

Every vehicle-behavior-instance must occur at a position.

Every vehicle-behavior-instance must occur at a time.

Every vehicle-behavior-type must be generated by one vehicle-type.

The following objects — signature-type, signature-instance, signature-perceived, target-type, target-instance, traffic-pattern-instance — and the facts related to them, while important, were not included in the model at this time because of the lack of an appropriate technical expert. However, this was not a major problem because the purpose of this model for this application was exploratory — i.e., to investigate the value of the methodology rather than to actually develop a working prototype, which was the purpose of the On-Site Inspection model described in another report, *Information Model for On-Site Inspection System* (SAND97-0049).

4. Data Fusion and Information Modeling Technology for Nonproliferation

This section is an expansion of a panel presentation originally prepared for the DOE Expo 93 for Intelligence and Special Operations at Oak Ridge. Although this part of the report is a higher-level, more general discussion of data fusion and proliferation issues, it is included because it was a logical result of the LDRD, Information Integration for Data Fusion.

Proliferation is a critical national security issue and is becoming even more important given the rapidly changing world situation. Information is an essential resource in addressing this issue. This part of the paper discusses two information integration technologies that can enhance our ability to collect and analyze the data needed to detect and respond to proliferant behavior in a number of areas — nuclear, chemical, biological, missile technology, etc. These two information technologies — data fusion and information modeling — are applicable to all types of proliferation. In fact, as this LDRD has shown, they are also applicable to many non-defense-related areas such as manufacturing control, health care, and environmental monitoring.

This section of the report reviews some of the proliferation and arms control concepts and issues, and explains how data fusion relates to proliferation.

Proliferation and Arms Control

Arms control is the more generic term, with proliferation being a special type of problem. Arms control agreements or treaties may be bilateral or multilateral (in a few cases countries may even take unilateral action). These agreements may involve one or more types of weapons (e.g., nuclear, chemical, or conventional) or simply the technology for a certain type of weapon (e.g., the Missile Technology Control Regime).

An arms control agreement may apply to one or more geographic areas, e.g., nuclear weapons-free regions such as Latin America.

Finally, these agreements cover one or more functions, such as design, development, testing, production, deployment, use, stockpiling, retirement, and destruction of the treaty limited items. However, it is important that any function being monitored or controlled provide a signature or trail that can be detected in some way — otherwise the agreement is not verifiable.

In the past, arms control work has focused on weapons of mass destruction — nuclear, chemical, and biological. This work is becoming even more serious with the spread of advanced delivery systems, which are beginning to be controlled with the MTCR (Missile Technology Control Regime). While weapons of mass destruction remain the main thrust of much of the arms control activity, there is an increasing interest in conventional arms control, especially given the CFE (Conventional Forces in Europe) agreement and the increasing ethnic violence and instability in many parts of the world.

In fact, some research is now addressing the even broader issue of proliferation of the industrial base needed to support an arms industry for certain types of weapons (either of mass destruction or sophisticated conventional weapons). This issue is of growing importance, given that the effective weapons used in the Gulf War were not designed and manufactured with current state-of-the-art equipment and processes, but with 10- to 15-year-old technology, which is now broadly available.

The information modeling work described in this paper is applicable to all of these areas. The information models and the database designs for data fusion would be very similar, although the actual data, sensors, and data collection mechanisms would be different depending on whether the target of interest is a Scud or a tank, and only slightly different if the target is a manufacturing process.

Countries can be related to arms control treaties in a dichotomy along either of two dimensions: For example, consider the NPT (Nuclear Proliferation Treaty) trying to control the spread of nuclear weapons. Every country has either signed the treaty or not, i.e., is or is not a party to the treaty. Similarly, but independently, every country either possesses or does not possess nuclear weapons. The four categories are shown in Figure 2. This same taxonomy applies for all of the treaties involving weapons of mass destruction. (This taxonomy is less useful with conventional weapons treaties because most countries possess most types of conventional weapons, and different types of conventional weapons are more easily interchangeable. In this case, only the first dimension applies, i.e., whether or not the country is a party to the treaty.)

		Signed Treaty	
		Yes	No
Has Capability	Yes		
	No		

Figure 2. Countries' participation in arms control treaties vs. their possession of nuclear weapons.

This taxonomy could be refined (see Figure 3) by distinguishing between whether or not a country possesses actual weapons or the technology to produce them. In the past, there were three distinct possibilities:

(1) A country has neither the technology nor the weapons;

(2) A country has the technology, but does not yet have the weapons; or

(3) A country has both the technology and the weapons.

In the past, a country had to develop the technology before it could develop the actual weapons using that technology. This was true because, in general, no country would deliberately transfer the technology and certainly it would neither transfer nor sell the actual weapons.

Today, however, as Iraq shows, the technology is clearly being transferred — in some cases inadvertently, but in other cases deliberately and covertly. Therefore, today a country may simply buy the necessary technology it needs rather than support a major development program. These purchases may be much harder to detect because they may leave a much smaller signature or trail of evidence.

With the breakup of the Soviet Union, there is also a concern that some actual weapons themselves may be sold or transferred to non-nuclear countries. This means that a country could acquire these weapons without having the underlying technology to develop them. This transfer of actual weapons should also leave a relatively small signature or trail for detection.

In addition to being harder to detect, these transfers of either technology or weapons both reduce the costs and shorten the time that a proliferant country needs to develop a serious capability with these weapons. The possibility of these types of transfers (of either technology or weapons) has a major impact on the data collection and analysis needed to determine whether or not a country is in compliance with a specific arms control treaty or whether it is attempting proliferant behavior.

		Has Technology	
		Yes	No
Has Weapons	Yes		
	No		

Figure 3. Countries' possession of technology to produce nuclear weapons vs. actual possession.

Compliance

Compliance issues involve parties to the treaty, whereas proliferation issues involve countries that do not yet possess a specific weapon technology or type of weapon. Compliance involves verifying that the parties to the treaty are conforming to the specified conditions. Verification protocols are explicitly specified in the treaty. It is important that the verification protocols provide adequate data collection and analysis procedures to ensure that significant violations of the conditions with respect to prohibited or treaty-limited items can be detected.

Proliferation issues involve actions by countries that do not possess a specific type of weapon or weapon technology to acquire those weapons or technologies. Proliferation issues also involve the countries, companies, or other organizations that supply those weapons or technologies. Many of the same verification methods and technologies (National Technical Means) are used to determine both compliance and proliferation, although some methods (such as on-site inspections) are only applicable to compliance because they are specified in the treaty and do not normally apply to those countries that are not parties to the treaty.

Arms Control Verification

Arms control verification (including data collection and analysis) for either compliance or proliferation involves looking for evidence of specific treaty-limited items. These items may be the weapons themselves (nuclear, chemical, biological, or in some cases, conventional weapons) or the equipment or processes (such as enrichment) to make or use these weapons. In some cases, the evidence may involve the trail left by the purchase, sale, payment for, or shipment of these items.

A key point is that rarely is a piece of evidence unambiguous (e.g., actually finding a nuclear weapon in a non-nuclear state) or independent of other evidence. Each piece of evidence simply increases or decreases the probability or confidence in an inference about proliferant behavior. For example, a piece of evidence from one point in an acquisition chain provides you with much less confidence than if you have independent pieces of evidence from each link in the chain.

Dual Use Items and Technologies

Dual use items and technologies create an additional complication. In addition to their weapon use, they also have legitimate commercial uses. In this case, intentions, which are harder to determine, become important. In other words, you are more likely to find dual use equipment, but it provides much less evidence or confidence of proliferant behavior.

This type of arms control analysis requires collecting, analyzing, and integrating data from many sources including satellite-, air-, and ground-based sensors, on-site inspections, and open literature and news reports.[1] In fact, open source intelligence is becoming an important area of data collection and analysis.

Verification Methods

In *Constraining Proliferation: The Contribution of Verification Synergies*, eight general methods of verification are identified and the synergies among them are discussed. These methods are:

- Technical Means, including
  - National Technical Means (NTM)
  - International Technical Means (ITM), and
  - Multilateral Technical Means (MTM)
- National Intelligence Means (NIM)
- Data Exchange
- Notification
- On-Site Inspections (OSI)
- Aerial Inspections
- Open Skies
- Confidence Building Measures (CBM)

Constraining Proliferation discusses various types of synergy in detail. Examples of synergy include:

(1) one method's findings may trigger another more precise method by suggesting when and where to look for something;

(2) a method may confirm or support information from another method; and

(3) one or more methods may provide enough information to calibrate or refine another method.

1. As an example of the split between these functions, one source estimates that in the 1970s in the U.S., 91 percent of the intelligence costs went to collection and only 9 percent went to all-source analysis.

The key point is that for this synergy to occur, there must be some overlap or commonality in the analysis. Unless the information models that support the various methods are compatible, the data cannot be integrated into a common analysis.

Also there is efficiency in reusing common information models across methods. For example, the image of the same target will look different to different sensor types or even to the same sensor type depending on whether the platform is a satellite or an aircraft flying under the Open Skies agreement. However, the model of the target's features, capabilities, and behaviors should be identical and therefore modeled only once and reused by all of the methods, although the features and behaviors may appear different to different types of sensors. Furthermore, much of the information model, characteristics, and management of the overall images and target images is similar, regardless of the sensor type or platform.

Data Fusion: Relationship to Proliferation

Data fusion research can benefit proliferation and verification work in two ways. First, generic information models (e.g., those involving targets, behaviors, environments, signatures, and sensors) can be customized to address the objects of interest for specific proliferation and verification areas such as nuclear, chemical, biological, or conventional. These objects may be tanks, missiles, test facilities, and production facilities. By using common or overlapping but consistent information models, data can be combined from various approaches to provide the type of synergy mentioned above. For example, satellite or aerial data may suggest the need for an on-site inspection or a permanent unattended ground sensor.

Second, this research can generate a generic information model to support the type of Bayesian statistics needed to infer what is happening, based on conditional probabilities when given data from many sources. In arms control, a piece of data is not simply a piece of data in isolation — its context is critical. The importance and meaning of a piece of information, regardless of how it is collected, varies depending on the other data that have been collected. This type of analysis, called Bayesian statistics, is important because in most cases what you see is only evidence that must be used to infer whether or not a country is in compliance with a treaty or whether or not it is trying to conceal some proliferant behavior. For example, the discovery of a nuclear weapon possessed by a non-nuclear state is clear evidence of proliferation. There is little to infer. However, the presence of a nuclear reactor and certain dual use equipment and technology does not necessarily mean proliferant behavior. It is simply evidence on which to base an inference. Other things being equal, the same evidence found in Iraq would be considered much more threatening than if it were found in Switzerland.

With Bayesian statistics or conditional probability, the analysis becomes more complicated, but more precise and valuable. For example, the presence of certain high-precision dual use machine tools alone may tell you very little about the likelihood that proliferant behavior is occurring. On the other hand, the presence of this equipment, given that you have already detected certain chemical residues in the factory, may be much stronger evidence of such behavior.

Bayesian statistics and the information model that supports the necessary conditional probabilities have two benefits:

- First, it improves the quality of the inferences you can draw from a certain set of evidence.

- Second, it can suggest at any point in an analysis (such as an inspection) what information or analysis should be determined next, given what has already been found. This is an important benefit, considering the limited data collection and analysis resources that are available, especially at critical times such as during an on-site inspection.

This Bayesian approach and the information model to support it are common across all inspection and arms control areas. It is equally applicable to nuclear, chemical, biological, and conventional arms control. However, the actual data that needs to be collected and the knowledge base are different for each area, and perhaps even for different subareas. Furthermore, in some cases the necessary conditional probabilities may be difficult to determine. This approach is also fundamental to the entire data fusion area.

Realize that having the information model is only part of the solution. You also need the actual data and the conditional probabilities, both of which may be difficult or impossible to obtain in some areas.

Intentionally Left Blank

5. Summary and Conclusions

This report has provided an information integration perspective on data fusion and has addressed several defense-related application areas. The LDRD, Information Integration for Data Fusion, considered an even broader perspective by also addressing areas such as manufacturing and health care.

The first section of the report provided a brief introduction and framework for data fusion and the information modeling approach that was used by this project.

- First, we explained why data fusion is important and the use of single and multiple sensors, types of sensors, and levels of intelligence being built into sensors.
- Second, we reviewed types of data fusion applications in terms of environments (designed, real, and hostile worlds) and their characteristics, and suggested that there is significant overlap in the application characteristics, which could provide opportunities for synergy across application types.
- Third, we briefly described the information modeling methodology used and identified a basic generic model with five common object types: targets, behaviors, environments, signatures, and sensors.

In the remaining sections, we discussed data fusion within a specific context and describe some of the related models. In Section 2, we discussed and reviewed a previously completed ATR/SAR model in the context of this more general work. In Section 3, we described the application of the information modeling methodology to a hypothetical truce management application and included a preliminary model that could be extended to support a working prototype system. In Section 4, we provided a high-level discussion of proliferation and arms control issues and described how data fusion could support greater synergy among various verification and inspection methods. Another Sandia report, *Information Model for On-Site Inspection System* (SAND97-0049), describes the model that was used for a prototype system to support offensive and defensive inspections, which are one aspect of arms control.

In summary, the LDRD confirmed the two hypotheses it was intended to test: First, the natural language-based information modeling methodology could be effectively used to define and model the requirements for various data fusion types of applications. Second, the methodology could help identify commonalities across application areas to achieve greater synergy. The project identified five common object types that are the basis for all of the data fusion areas considered: targets, behaviors, environments, signatures, and sensors. Many of these objects and the facts related to them were common across multiple application areas and could easily be reused. In some cases, even the terminology was the same. In other cases, different areas had their own terminology, but the concepts were the same. In either case, there could be synergy across the areas.

Intentionally Left Blank

6. References

Olin H. Bray, *Computer Integrated Manufacturing: The Data Management Strategy*. Digital Press., Bedford, MA, 1988.

David A. Dornfeld, "Sensor Fusion," in Nello Zuech, ed., *Handbook of Intelligent Sensors for Industrial Automation*. Addison-Wesley Publishing Co., New York, 1992.

Patricia Bliss McFate and Sidney N. Graybeal, *Constraining Proliferation: The Contribution of Verification Synergies*. External Affairs and International Trade, Ottawa, Canada, 1993.

J. J. V. R. Wintraecken, *The NIAM Information Analysis Method: Theory and Practice*. Kluwer Academic Publishers, Norwell, MA., 1990.

John Zachman, "A Framework for Information Systems Architecture," in *IBM Systems Journal*, vol. 26, No. 3. September 1987.

Nello Zuech, ed., *Handbook of Intelligent Sensors for Industrial Automation*. Addison-Wesley Publishing Co., New York, 1992.

DISTRIBUTION:

1	MS 0160	Virgil Dugan, 4500
1	0188	Chuck Meyers, 4523
1	0459	John Taylor, 5335
1	0473	Sharon Deland, 5391
1	0661	Gary Rivord, 4012
1	0863	Paul Flores, 14307
1	1138	Larry Claussen, 6533
1	1373	Arian Pregenzer, 5341
11	1378	Olin Bray, 4524
1	1378	Shanna Narath, 4524
1	9018	Central Technical Files, 8940-2
5	0899	Technical Library, 4414
2	0619	Review & Approval Desk, 12690 For DOE/OSTI