Satellite Data Exploitation
Design Architecture

Lance Self
Air Force Research Laboratory
Space Vehicles Directorate
Kirtland AFB, NM
Lance.Self@kirtland.af.mil

Abstract

The data exploitation system described within is being designed to take advantage of the existence of terabytes (or more) of data sets that have been sent down from various Air Force satellites that can be used to detect previously unknown anomalies that occurred while the satellites were operational. This wealth of information can then be used and exploited to build more reliable and robust future generations of satellites. And, because various types of data sets are sent to the ground consisting of satellite state-of-health, satellite system telemetry, as well as on-board experiments, threats to these satellites will also be able to be detected, given that the satellites have the appropriate sensors. Designing a data exploitation capability is an important development area for realizing anomaly and threat detection objectives. This paper discusses a framework to meet those objectives and is being designed for integration within the Air Force Research Laboratory modeling & simulation environment.

Introduction

One area of research at the Air Force Research Laboratory (AFRL), Space Vehicles Directorate falls under the general heading of Counter Space, both offensive and defensive, and Situational Awareness; situation awareness being a direct resultant of counter space. Four areas that form the basis of counter space, and therefore any countermeasures that may take place as a result of a true threat to a satellite, are data mining, fusion, analysis, and some level of decision support (Figure 1). These four areas are the "tool sets" with which analysts use to (a) detect an anomaly or threat (b) positively identify it (c) determine its location or point of origin and (d) classify the perceived anomaly or threat. If it is a real threat or an actual anomaly then appropriate countermeasures can be initiated with some degree of confidence.

AFRL is designing and developing a system that will fuse, mine, analyze, and eventually hand off to a decision support system, satellite data sets to be used to detect previously unknown anomalies and threats to satellite systems. The system will be ground based and be integrated within the modeling & simulation development environment. Development is in the early stages but eventually this exploitation system will be fully integrated and capable of exploiting satellite system telemetry, experimental, and threat data sets.

This paper describes the architecture of the exploitation system and the two of the four main subsystems: data fusion and data mining. Two objectives of the data exploitation design is to create a system that combines fusion, mining, analysis, and decision support into a single cohesive system so eventually customers can "plug-n-play" software modules that meet their requirements and can do so in an integrated environment.

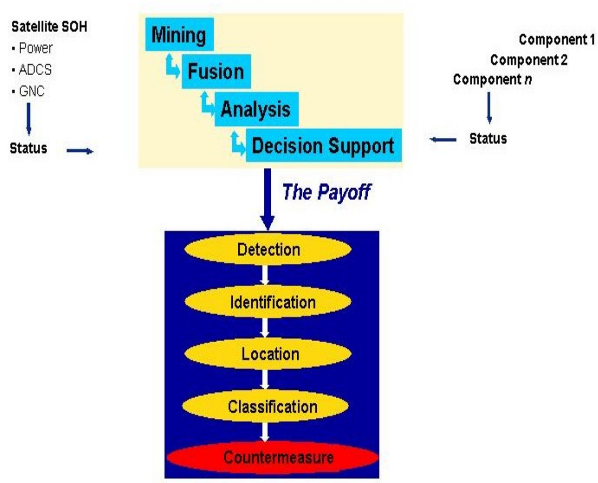

Figure 1:

Pyramid Design

Forming the base of the design pyramid (Figure 2) at the engineering level are the four main data exploitation subsystems: data fusion, data mining, data analysis, and decision support. Each of these subsystems contribute to the overall knowledge that forms the pyramid and feed the levels above it.

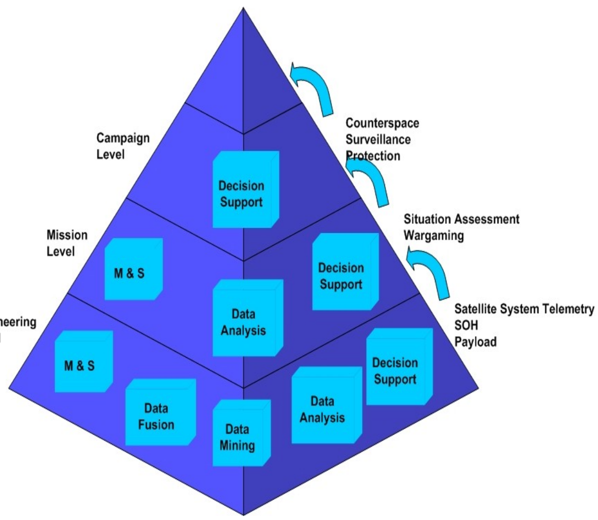

Figure 2: Pyramid Design

At the various levels of the pyramid the users of the data sets, and the information they need, change accordingly. The pyramid is designed to aggregate existing data, and merge new data sets, to meet higher and higher levels of user information requirements.

At the lowest level of the pyramid, the engineering level, data sets consist of satellite system and sub-system telemetry, satellite state-of-health, and payload data sets. These are the lowest form of data and these data sets are mined and fused to form the basis for the next higher level of knowledge in the pyramid. At the engineering level the users of these fused and mined data sets are primarily satellite engineers and satellite system analysts.

At the mission level further refinement of the data sets are performed. In addition to data analysis and decision support functions is modeling & simulation. Modeling & simulation is needed at all levels of the pyramid. At the mission level, however, modeling & simulation is especially relevant to the customers who are concerned with situation assessment. Lower level data sets are fused with existing higher level data sets that give the user a "big picture" for assessing the situation of the individual satellite or satellite cluster. These data sets may also be useful for users primarily concerned with wargaming activities.

At the highest level of the pyramid is the campaign level. At this level users of the fused and mined data sets integrate other data sets from other government agencies and is used to make high-level decisions such as offensive, defensive, and countermeasure actions. New paradigms are being considered that allows individual satellites, and clusters of satellites, to defend themselves against attacks.

Data Exploitation

At the center of the data exploitation system is the Distributed Architecture Simulation Laboratory (DASL) that hosts the various tool sets and can simulate an operational environment. This laboratory (Figure 3) is internal to the Space Vehicles Directorate and houses all the necessary base components such as modeling & simulation hardware and software, telemetry collection hardware and software, a commercial relational database management system, and an internal network for distributed processing.

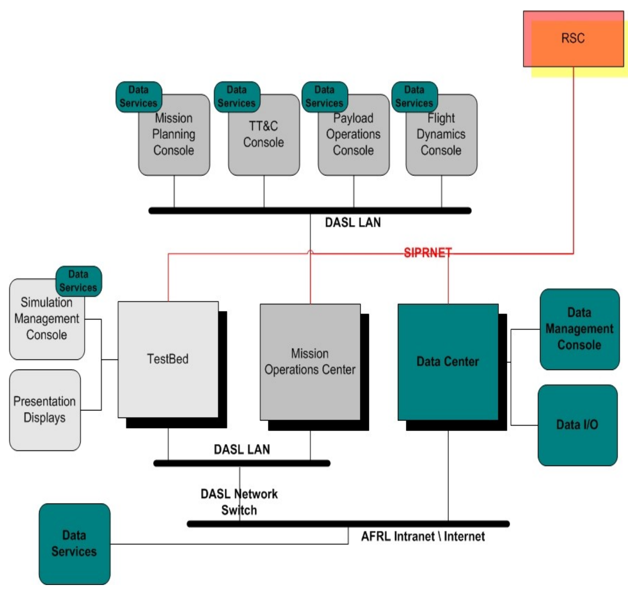

Figure 3: Distributed Architecture Simulation Lab (DASL)

The data exploitation architecture drawing (Figure 4) defines the boundaries for this system and the exploitation system resides fully within the DASL.

As previously mentioned, the data exploitation system consists of four functional areas: data fusion, data mining, data analysis, and decision support. In Figure 4 each of the four functional areas is logically represented by a workstation but they could be physically implemented on a single, multi-processor workstation.

The DASL data exchange network provides the link between:

- The data exploitation workstation and the rest of the DASL components (M&S, mission operations center, GN&C (guidance navigation & control), wargaming, data center) including access to the Internet for data archive and retrieval
- The exploitation workstation and supporting databases

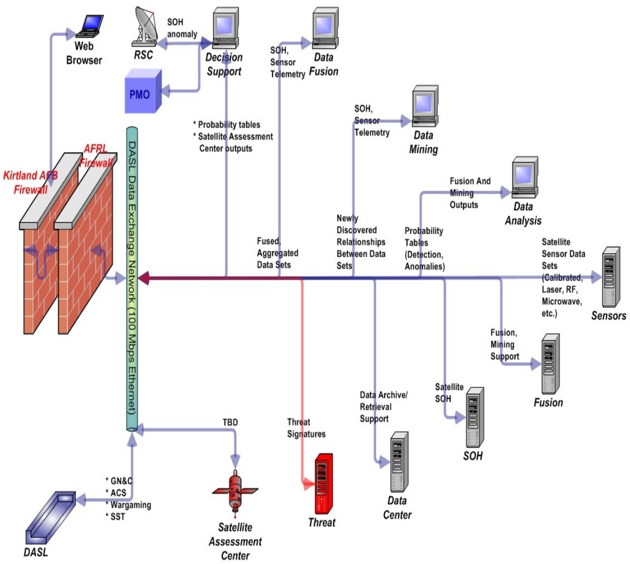

Figure 4: Data Exploitation Architecture Design

The Data Exploitation subsystem must be capable of handling hundreds of gigabytes of data flow across the network and internal processing. For example, in one month an AFRL satellite, sent 240 million records of satellite telemetry data to the ground; equivalent to 9.4GB. This data volume, coupled with the resultant intermediate data sets generated by the various tool sets, will demand a high data rate internal network.

Data flows between each of the functional areas (refer to Figure 4). Output from one area may be input to another area. For instance, outputs from the mining area (i.e., new relationships between power and guidance satellite subsystems) may feed into the fusion function area to train a neural network. Those data sets may, in turn, flow back for more mining. It is the intent of the exploitation system that data mining and data fusion each become "pre-processing" for the other and that the flow of data is controlled, in large part, by the analyst.

Resultant data sets from the mining and fusion processes flow into the analysis functional area. It is yet to be defined what these resultant data sets will consist of, their format, etc. Analysts have, in the past, taken fused and mined data sets and plotted the outputs searching for patterns. This method requires a significant amount of experience with the telemetry data as well as knowledge of the internals of the satellite being analyzed. One of the goals of the data exploitation system is to make it easier for analysts to choose their analytical methods thereby fostering a more thorough and complete final satellite system analysis.

Outputs from analysis flow into the decision support function area and outputs from decision support flow to the RSC. The RSC is the main telemetry receiving station for AFRL satellites and has the responsibility for the health and safety of the satellite. It conducts its own analysis but is usually limited to those specific health and safety factors. At a minimum exploitation data sets can be used as a "sanity check" against the RSC data sets. For example, did the exploitation system detect a satellite state of health (SOH) anomaly that existed or was it a false detection suggesting the fusion/mining subsystems need calibration and more neural network training?

Data Fusion / Data Mining Functions

The objective of the data fusion functional area subsystem is to fuse satellite system telemetry and satellite experiments. This includes fusing homogeneous (SOH/SOH) data sets as well as heterogeneous (SOH/sensor) data sets.

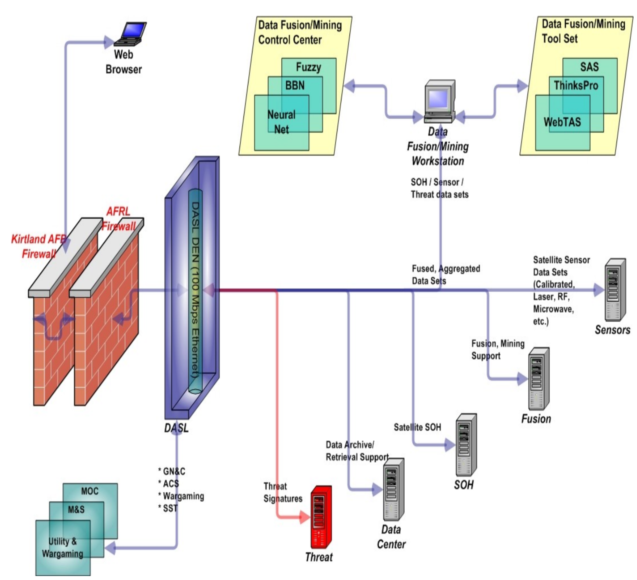

Figure 5: Data Fusion / Mining Architecture Design

The Air Force Research Laboratory has four Small Business Innovative Research (SBIR) efforts that are concerned primarily with data fusion and secondarily with data mining. In addition the laboratory has commercial fusion software and database integration software at its disposal.

Although differing in purpose and objective the data fusion and data mining functions share a common architecture. As Figure 5 depicts the data fusion/mining workstation receives satellite system telemetry and payload data sets. At the workstation the analyst decides to fuse the data or mine the data. What the analyst decides to do is dependent upon their experience level and objective, however the analyst has complete control from this single workstation of what to do next. Also at his/her disposal are the tools necessary to complete the analysis.

As mentioned previously AFRL has funded research into the fusion area and will receive comprehensive fusion algorithms from these researchers. Their work will be integrated into, and form the basis of, the fusion/mining workstation. At times, for example, a Bayesian approach may be more suitable then a neural network or fuzzy logic approach. All these methods will be available and "on-call" to the analyst.

The purpose of the data mining subsystem is to discover previously unknown relationships among data sets and discovering patterns within data sets. The resultant data set is used to "predict" data points. Statistical inference techniques such as data trees and data clustering, link analysis, pattern recognition / analysis / detection software are some of the data mining tools at the analysts disposal. As expected some of the same tools used in fusion are also used for mining.

Some key areas that are critical to the success of the creation of the Fusion/Mining Control Center are:

- Forming a consensus regarding data format(s)
- Reformatting data sets into a standard format(s)
- Fusion processes being able to be instantiated by the Fusion/Mining Control Center application
- Writing user friendly GUI's to control the process

Conclusion

This paper described a data exploitation system that is meant to take advantage of the existence of vast volumes of satellite data sets. It discussed an architecture design that will enable satellite engineers to analyze satellite systems and subsystems from a single workstation with all available tools at their disposal.

Designing a data exploitation capability is an important development area for realizing satellite experiment objectives. It brings together, in a single cohesive unit, the tools to learn from past mistakes, use of modeling & simulation to design faster more efficient and robust subsystems, and provides a framework for testing existing and future satellite requirements.

Acknowledgement

The author would like to acknowledge the research efforts being conducted in these areas by Dr. Chris Tschan and Dr. Christopher Bowman.