White Paper
Information Exchange Requirements (IER) Driven Fusion

2013-08-03

1. Technical Concept

The technical concept is shown in Figure 1. It has four principal elements:

a. Information Exchange Requirements Processor (IER-P) that decomposes IERs to inter-related object and events and then links them to types of sensor and source support evidence. This is necessary since IERs are usually not directly observable but are, rather, satisfied by fusion of multiple sensors and sources. The IER ontology dictates the workflow.

b. BrainLike Process (BLP) that tailors FMV and imagery feature extraction to provide the required evidence

c. Sources Query Process that prepares Hadoop map jobs to retrieve object and event of interest data from DCGS-N sources

d. Fusion Process that performs, 1) Hadoop reduction using the returned DCGS-N key-value pairs, and 2) updates likelihoods in realtime as sensor features arrive.

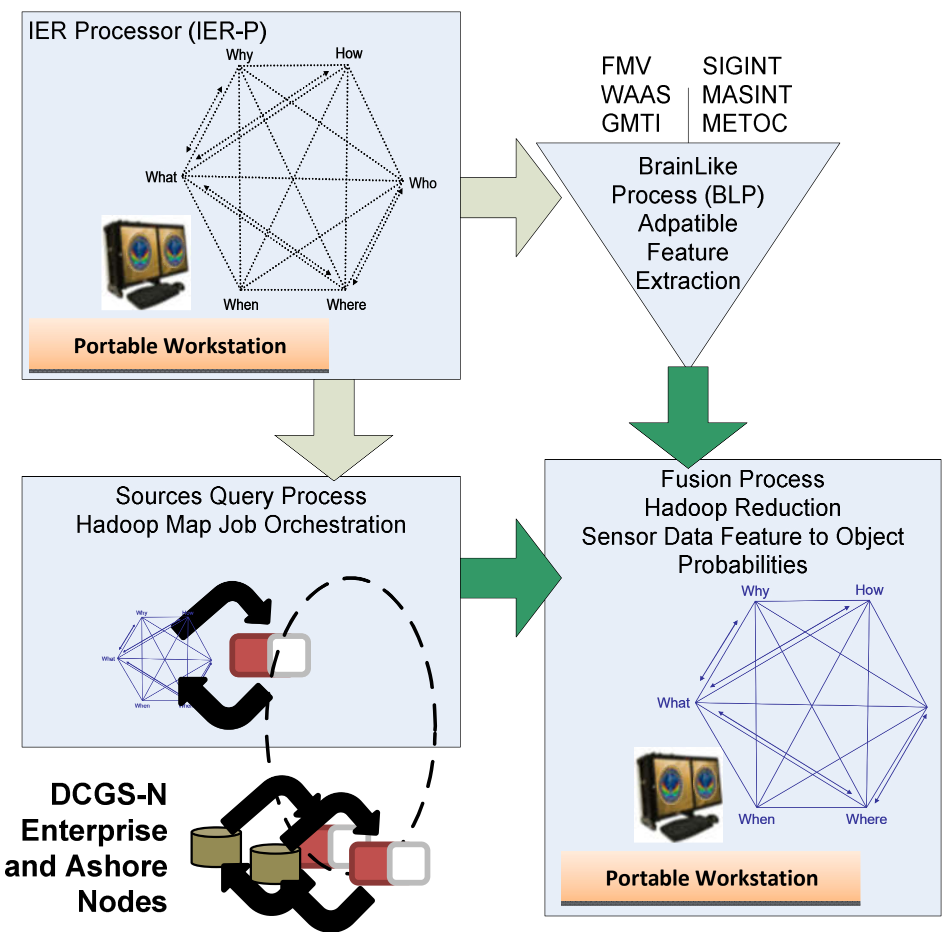


Figure 1. Technical Concept

The issues to be resolved are:

a. IER content model

b. Model of IERs to sensor feature and source data types

c. Tailorable feature extraction

d. DCGS distributed source queries

e. Fusion of sensor and source data

1.1 Technical approach proposed to resolve these issues

The IERs are modeled in simple ontology tools developed by Silver Bullet and using a IER object model Silver Bullet developed for the DoD CIO and in widespread use throughout DoD. This model is founded on a formal set-theoretic and 4-dimensional mereotopologic foundation and models most aspects of the six principal interrogatives. It also has representation and causality patterns that can link sensor feature types and sources to objects and events. It is compatible with the warfighting domain of NIEM which is largely based on C2 Core which, in turn, was derived from JC3IEDM.

The causal or evidential links of these IER objects to sensor and source data types involves specializations of the p(evidence | object, sensor / source). In the Joint and Operational Track Manager, this is the link from Observations to Object.

For sensor data, the BLP is instructed to extract features of interest in a specific time period and area. Data gets tagged with priority level by Brainlike. If the BLP is located at the sensor, the communications bandwidth required is reduced over sending raw sensor data since only extracted features and of-interest video snippets are sent.

Since source data can be voluminous and is usually distributed across the DCGS nodes, the query is formulated a Hadoop MapReduce job which leaves the data at the sources and instead provides job instructions to the nodes. The nodes return key-value pairs instead of full datasets to reduce communications load. The ontology for the IERs and sensor and souce linkage is a highly normalized, meaning there is one value per key so there is no translation required to go to or from Hadoop key-value pairs. At the DCGS-N nodes, many of the key-value pairs are based on DDMS metacards. Silver Bullet developed a mapping from the IER model to the "Summary Content" part of the DDMS schema which will be used to translate between the IER ontology and the DDMS schema. As the sensor data features are extracted, Hadoop stream processing is used to update the fusion probabilities for the IER objects.

The workflow is dictated by the links in the IER ontology which trigger processing appropriate for the new or retrieved information .

The IER processor includes a comprehensive pedigree model and capability. Pedigree is maintained using a workflow model developed by Silver Bullet and made available in key-value pairs. A pedigree model we use is similar to the Open Provenance Model (OPM) [1] and that was recommended in the Warfighter Information Processing Cycle (WIPC) concept paper by PEO-C4I [2] but we have since conformed it to a formal upper level ontology and to a general workflow model. Pedigree and source metadata is also used for user verification of belief, including access to data related to metadata, and detection and filtering of redundant data usage.

1.2 Assessment of the proposed new capability over the existing state of the art

Existing	**Proposed New**
IER satisfaction is pre-determined at system development time	IER construction using IER model developed for DoD CIO

IER selection is directly related to sensors and sources	IER decomposition to objects. An object type to sensor and source feature type model.
Feature extraction is pre-determined at sensor processor development time	BrainLike Process interface to IER processor to tailor the feature detector and extractor
Feature extraction is performed at the processing node, not at the sensor	BrainLike Process is designed to be hosted at sensor
Source queries require movement of data from sources to query node	Hadoop MapReduce jobs send the query to be performed to the nodes
Probabilistic data fusion for complex IERs has a high computing demand	Hadoop MapReduce jobs send parts of the fusion algorithm to be performed to the nodes

1.3 Comparison against competing technological developments

a. Semantics for machine understanding of "content". Competing semantics are usually only set-theoretic. 4-D mereotopology and representation patterns add additional semantics important for this problem set.

b. Mature machine understanding of information product needs. This is usually based on a direct relationship between needs and sources. But most user information needs are the result of fusion derivation from sensors and sources. The decompostion of IERs into their component objects and to causal links of sensor and source datatypes to object types, enabled by the ontology, would advance this understanding. The IR-object model and the sensor queuing and source queries use the same evidence-object and object-IR model so that queing and querying are the inverse of fusion.

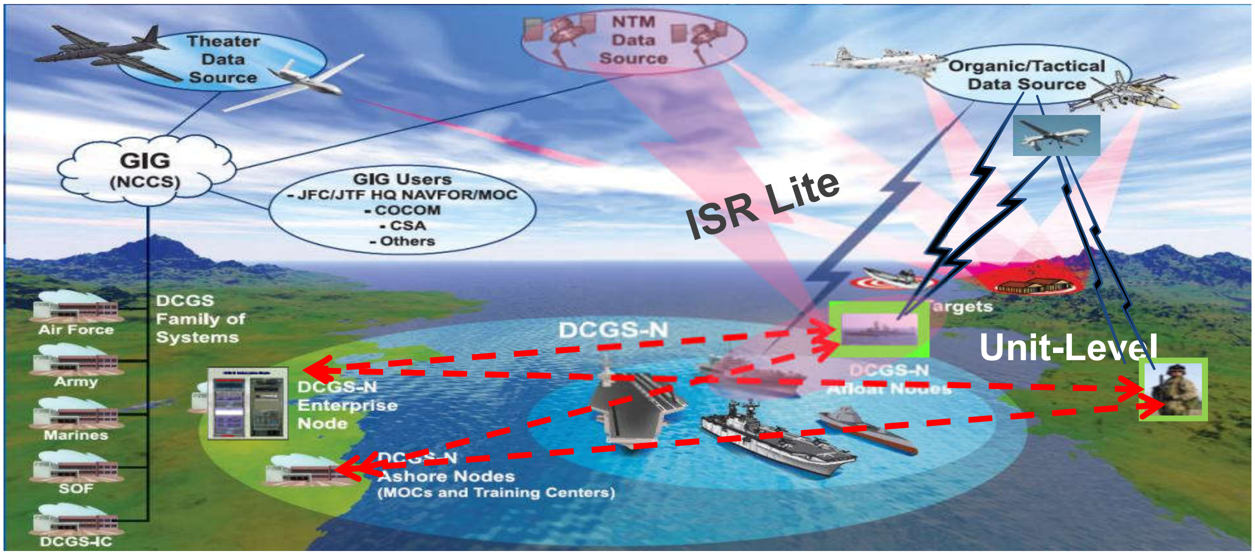
c. Workflow managers for fusion. These are not usually directly based on a causal ontology, i.e., the IER to sensor and source evidence model. Pedigree as workflow would enable information products with attached pedigree as a result of recording the orchestration of sensor queuing, source MapReduce jobs, and sensor updates to the IER belief states.

d. Creation of user defined tags to customize IERs for specific mission sets. Competing technologies lacks formality. The proposed IR Processor uses typeInstance and superSubtype to extend IER templates for specific mission sets.

e. Fusion of information to meet needs of specific knowledge domains. Often data and information fusion are sensor and source driven, not requirements driven. Even the JDL levels suggest the starting point is the sensor.

2. Operational Naval Concept

Small unit operator receives commander's IRs, formulates using IER tool and ontology, BLP at sensors is tailored, MapReduce jobs orchestrate source jobs, partial fusion results are aggregated at unit



Bibliography

1 Luc Moreau (Editor), Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Beth Plale, Yogesh Simmhan, Eric Stephan, Jan Van den Bussche; The Open Provenance Model Core Specification

2 *Pedigree Standard Processes Concept Document*, Version 0.0; PMW-150, PEO C4I; Feb 2008