Towards Hard+Soft Data Fusion: Processing Architecture and Implementation for the Joint Fusion and Analysis of Hard and Soft Intelligence Data

Geoff A. Gross, Rakesh Nagi, Kedar Sambhoos, Daniel R. Schlegel, Stuart C. Shapiro, Gregory Tauer
Center for Multisource Information Fusion
State University of New York at Buffalo
Buffalo, New York, U.S.A.
gagross@buffalo.edu, nagi@buffalo.edu, sambhoos@cubrc.org, drschleg@buffalo.edu, shapiro@buffalo.edu, gregoryt@buffalo.edu

Abstract— **Historically, data fusion has focused on processing hard or physical sensor data while soft or human observed data has been neglected within fusion processes. This human observed data has much to offer towards obtaining comprehensive situational awareness, particularly in a domain such as intelligence analysis where subtle connections and interactions are difficult to observe with physical sensors. This paper describes the processing architecture designed and implemented for the fusion of hard and soft data in the multi-university research initiative on network-based hard and soft information fusion. The processing elements designed to successfully fuse and reason over the hard and soft data include the natural language processing elements to form propositional graphs from linguistic observations, conversion of the propositional graphs to attributed graphical form, alignment and tagging of the uncertainties extant in the human observations, conversion of hard data tracks to a graphical format, association of entities and relations in observational hard and soft data graphs and the matching of situations of interest to the cumulative data or evidential graph. To illustrate these processing elements within the integrated processing architecture a small synthetic data set entitled the bomber buster scenario is utilized, presenting examples of each processing element along the processing flow. The value of fusing hard and soft information is illustrated by demonstrating that individually, neither hard nor soft information could provide the situation estimate.**

Keywords – hard and soft data fusion, information fusion, fusion architecture

I. INTRODUCTION

The multi-university research initiative on network-based hard+soft information fusion, [1], was developed to integrate the fusion of hard or physical sensor data and soft or human observed data. This project is probably the first comprehensive effort which attempts the integration of these data within a common framework. While hard sensor fusion is historically well studied within the fusion community, soft or human reported data is a recent data modality to be processed by fusion systems. Soft data has much to offer to the fusion process. These gains include the ability to more reliably measure certain attributes of interest (e.g., emotional disposition), simultaneously sense multiple attribute types (e.g., age and gender) and actively extract additional information (e.g., in an interview scenario). The successful integration of soft data into a fusion system requires some different processing and representational methodologies than traditionally utilized for its hard data counterpart. The difficulties in processing linguistic data are largely due to the many ways in which the same information can be represented. The unique characteristics of soft data which necessitate original processing methodologies include sloppiness and ambiguity in natural language, qualitative uncertainties extant in human observation and difficulties in evaluating similarities in non-standard data.

This research effort is focused on the domain of intelligence analysis, specifically within a counter-insurgency (COIN) operation. In this domain many sources of data must be simultaneously considered by intelligence analysts in an attempt to obtain situational awareness. Information requirements (IR) are provided in prioritized form (priority information requirements, PIR) to the intelligence analysts [2]. These IR are defined as “items of information which regarding the enemy and his environment which need to be collected and processed in order to meet the intelligence requirements of a commander” [3]. It is the job of the intelligence analyst to determine the degree at which IR fulfilling data has already been observed, what is conflicting and what remaining IR fulfilling data must be collected. The volume of current and historical data which must be considered to assess IR quickly overload manual data consideration. To overcome the intelligence analyst’s cognitive limitations automated methodologies must be developed. The automated fusion approach designed for this effort implements many of the JDL Fusion Processes [4]. Specifically, processing elements are designed for the source characterization of soft data, common referencing/alignment and data association within and across data modalities and situational assessment via a graph matching estimation process. In addition to these processes, the soft data processing stream requires natural language processing steps to convert natural language into a well-defined formal language.

The initial architectural design for processing of the hard and soft data modalities provided processing pipelines separated by data modality, fusing end products of these pipelines [5]. This decision was made based on the initial relative ignorance as to how soft data fusion should be performed to the more mature field of hard data fusion. The architecture presented here moves toward an earlier integration of hard and soft data. The earlier integration is preferable from the information-theoretic mindset that increased processing compounds errors in the data. The benefit of this earlier hard+soft integration as well as functionality of each of the fusion system processing elements are demonstrated with the help of the bomber buster scenario (BBS) synthetic data set which is described in Section III.

The remainder of this paper is organized as follows: Section II provides an overview of the fusion processing architecture, Section III discusses the dataset used to demonstrate the fusion processing architecture, Section IV describes the natural language processing (NLP) fusion tasks, Section V explains the graph conversion process from propositional to attributed form, Section VI explains the uncertainty alignment process, Section VII explains the conversion of hard data tracks to an attributed graphical format, Section VIII explains the scoring procedure for data association, Section IX describes the data association process, Section X illustrates the graph matching process and Section XI presents conclusions and plans for future work. Each section utilizes examples from the BBS to demonstrate important concepts within that section's processing elements.

II. Fusion Architecture Overview

The architecture developed for the joint fusion of the hard and soft data is pictured in Figure 1. The processing flow begins with the receipt of plain text and raw sensor reports for the soft and hard data processing streams respectively. The focus of this paper is on the soft processing stream. Details of the fusion algorithms utilized to produce the hard data tracks (which are consumed at the hard data track conversion step) are omitted here. The natural language processing (NLP) tasks of dependency parsing, recognizing and co-referencing entities, mapping syntax to semantics and contextual enhancement produce an ontologically enhanced propositional graph. The propositional graph is a directed acyclic graph (DAG) in which the nodes represent entities, acts, events, and relationships in the domain.

Propositional graphs are next converted to attributed graphs. Within the attributed graph, nodes represent observed entities (e.g., people, vehicles, locations) and edges represent relationships between these entities (e.g., family relationships, communications). Attributes are the finest grain observation within the attributed graph, representing observed characteristics of the entities and relationships for nodes and edges respectively. Type information is provided within the graph at both the node or edge level and at the attribute level. The provided types are hierarchically defined, providing a mechanism for filtering dissimilar entities, relationships and attributes in the downstream processing tasks of scoring, data association and graph matching.

At this point in the processing stream these attribute values are represented exactly as observed in the soft messages. The subsequent processing step of uncertainty alignment considers the relevant meta-data which is preserved through upstream processing to adjust for observational biases, variance and qualitative linguistic descriptors. For attributes that are numerically defined, numerical uncertainty representations are provided in the uncertainty alignment output. Attributes which are not numerically defined are either left as is or converted to a linguistic fuzzy set representation if appropriate.

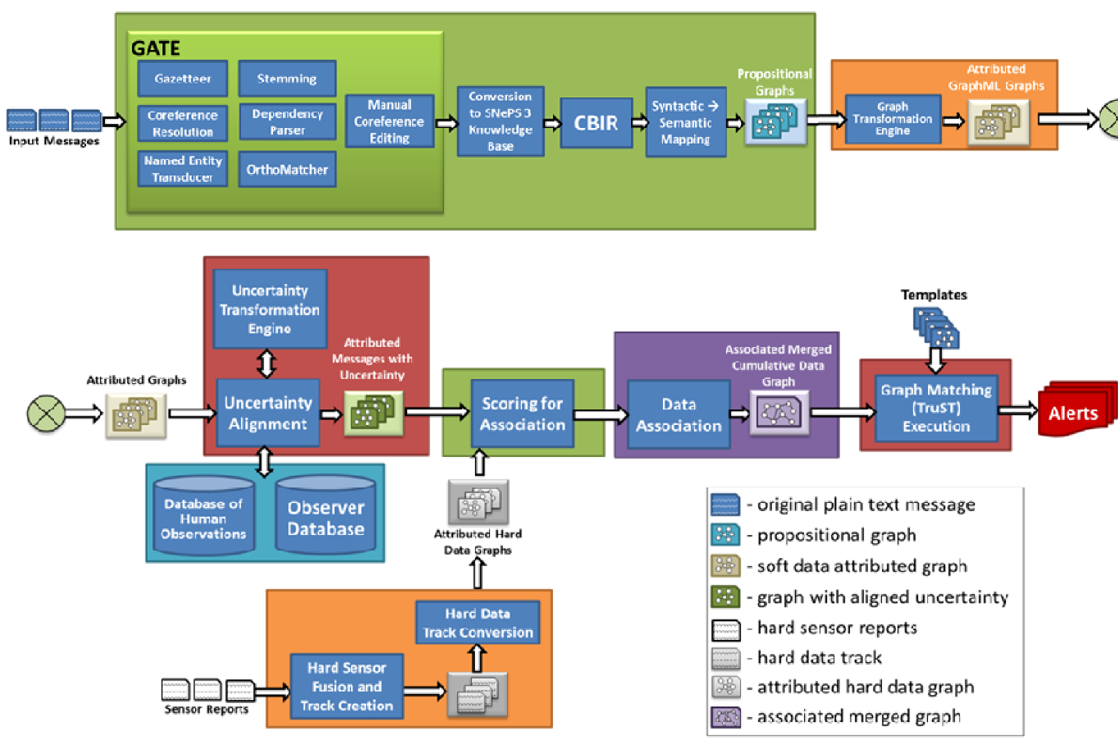

Figure 1. Integrated Hard+Soft Information Fusion Framework.

The next processing step is scoring for data association. At this stage the hard and soft data converge to be processed concurrently in the remaining processes. The soft data is represented in an attributed graphical format with uncertainty and the hard data is provided in a similar input format. The hard data is transformed into this format through a hard data track conversion process. This conversion process transforms the hard data tracks provided as the hard data fusion output into the appropriate attributed graphical format. Attributes provided in the track data such as entity color, dimensions and type are transferred to the scoring (similarity measurement) operation.

Scoring for association provides similarity values for observed data to the association process. These similarity values are used to disambiguate common entities across multiple observations and data modalities. The output of the association process is the cumulative data graph, which ideally represents each of the unique observed entities and relationships with a single node or edge respectively. Common overlapping attributes of graph elements are merged in the association process, leaving a single value for each observed attribute.

The situation awareness process within this architecture is a graph matching process. The graph matching algorithm attempts to locate within the cumulative data graph a situation of interest or template graph (e.g., a bombing plot which is in progress). The identification of a significant portion of a situation of interest within the cumulative data graph causes an alert to be provided to an intelligence analyst who can help the commander determine the appropriate course of action. The graphical format of the template graph is the same as the cumulative data graph, with attribute uncertainties provided where the analyst is uncertain of the value or would like results over a range for that attribute.

III. EXAMPLE SCENARIO

To better illustrate the processing elements within the fusion architecture a small synthetic scenario entitled the Bomber Buster Scenario (BBS) is utilized. This scenario consists of 12 messages (labeled BBS-#), 7 of which are human observed soft data reports and 5 which are hard data reports. Within the scenario observations are made by a variety of sources. Soft reports are provided by a local newspaper, a police tip-line, intercepted cell phone communications and a security team. The physical sensors which provide the hard data within the scenario include pan tilt zoom (PTZ) cameras, a light detection and ranging (LIDAR) sensor and an acoustic sensor. The message content is provided in Appendix A – Bomber Buster Scenario Messages. This dataset was designed to illustrate the synergistic effect of the fusion of both hard and soft data with neither data modality alone providing the required observations to "bust the bomber."

IV. NATURAL LANGUAGE PROCESSING

Tractor [6] is a set of subsystems that perform Natural Language Understanding (NLU) on a single message at a time. The goal of Tractor is to input an English message, and produce a semantic propositional graph [7] representing the information in the message. A propositional graph is a knowledge representation (KR) formalism in which nodes represent entities in the domain, including propositions, and labeled directed edges represent "non-conceptual" binary relations between two entities. For example, consider the information that *Mahmoud Azhour is a person*. We represent this, using SNePS 3, the latest version of the SNePS KR family [8] as the two logical propositions, (Isa n173 Person) and (hasName n173 |Mahmoud Azhour|[1]).

That is, the logical constant n173 denotes an instance of the class Person who is named "Mahmoud Azhour". We visualize this as a graph in which n173, Person, and |Mahmoud Azhour| are nodes, a directed arc labeled "Isa" connects n173 to Person, and a directed arc labeled "hasName" connects n173 to |Mahmoud Azhour|. The benefit of a graph representation is that all information about some entity is reachable in constant time from the single node that represents that entity. Due to space considerations, the representation of propositions as nodes will not be discussed in this paper, but see [9] for a more complete discussion. The graphs Tractor produces are "semantic" because the terms denote entities in the domain rather than words or other strings of text, and the propositions denote meaningful, domain-level, conceptual relations.

Tractor operates in two major phases. In the first, natural language processing (NLP) techniques are applied to the English message to produce a syntactic graph in which most of the nodes denote textual items, and most of the relations are syntactic. In the second phase, syntax-semantics mapping rules are applied that transform the syntactic graph into a semantic graph. At this stage of our work, this transformation is not complete. We have been concentrating on producing conceptual descriptions of entities mentioned in the messages, and have not yet turned our attention to the description of actions and events.

A major assumption we are making about the messages is that a single message is written by a single person at a single time, whereas different messages might be written by different people at different times, without the author of one message being aware of the contents of previous messages. Therefore, NLP techniques are appropriate for intra-message coreference resolution, whereas inter-message coreference resolution must be based on semantic descriptions of the various entities, and is done by Data Association (see Section IX).

For the first NLP phase of Tractor we use GATE, the General Architecture for Text Engineering [10], which is a framework for plugging in a sequence of "processing resources"(PRs). We currently use eleven such PRs, most of which come from the ANNIE (a Nearly-New Information Extraction System) suite [11]. We will discuss only the most interesting of these eleven PRs. The ANNIE "Gazetteer" is a list-based named entity (NE) recognizer. In BBS-2, it identifies "Baghdad" as a city, "van" (three times) as a vehicle, "Mahmoud Azhour" and "Sayed Azhour" as the full names of people, and "tan" and "white" as colors. The ANNIE NE Transducer supplements the Gazetteer with rule-based named entity recognition using rules written in JAPE

[1] The "|" characters are escape brackets that make the enclosed material one symbol, including the blank.

(Java Annotation Patterns Engine) [10]. In BBS-2, the NE Transducer identifies “man” (twice in BBS-2 and twice in BBS-12) as type Person and gender male. We use a variant of the Snowball English Stemmer [12], as written by Paul Bunter. Bunter’s version uses a lexicon to make sure that the stem found for a word is actually a word in the lexicon. For example, whereas the stemmer supplied with GATE produces “recogn” as the stem of “recognized”, Bunter’s version correctly produces “recognize”. The ANNIE OrthoMatcher recognizes that some named entities refer to the same entity. For example it recognizes that “Azam Al-Azhar” and “Al-Azhar” in BBS-1 refer to the same person. The ANNIE Pronominal Coreferencer performs anaphora resolution using JAPE rules. It recognizes that “he” and “his” in BBS-1 also refer to Azam Al-Azhar. Finally, we use the plugin for the Stanford Dependency Parser [13] to produce a dependency parse of the message.

These eleven PRs run successively and automatically. However, the OrthoMatcher and Pronominal Coreferencer do not find all of the coreferring expressions, and some that they find are not correct. GATE’s Co-reference Editor provides a GUI for a person to examine the text of each message and the “coreference chains” that have been identified in each, to modify these chains, and to add others. The final result of our use of GATE is, for each message, a set of “annotations,” each consisting of an ID number, a Type, a starting and ending position in the sequence of characters of the message, and a set of feature-value pairs. The annotations produced by the GATE PRs are converted to a SNePS 3 knowledge base, which is simultaneously a set of logical expressions and a graph whose contents are the syntactic information represented by the annotations.

The second phase of Tractor maps the graph of mostly syntactic information (the exception being the information about types, which is semantic) to a semantic graph. This involves a change in the denotation of some of the terms, most notably the tokens. In the syntactic graph, a token denotes the occurrence of a word or text string. In the semantic graph, it denotes an entity in the domain. Consider the phrase “a white van” from BBS-2. The information in the syntactic graph derived from this phrase includes: (TextOf a n121), (TextOf white n153), (TextOf van n155), (det n155 n121), (amod n155 n153), and (Isa n153 color) (see the left hand graph of Figure 2). That is, n155 is an occurrence of the word “van”, modified by a determiner that is an occurrence of “a”, and by an adjective that is an occurrence of the word “white”, which was determined by the Gazetteer to be of type color. In the semantic graph, this becomes (Isa n155 van) and (color n155 white) (see the right hand graph of Figure 2). That is, n155 is an instance of the category van, and its color is white.

CBIR (Context-Based Information Retrieval), [6], adds background information that is relevant to the message information. Currently, CBIR’s main contribution is to add to the graph, for each noun in the message, the entire OpenCyc [14] ontological hierarchy above the noun. In processing the seven soft messages, 78 different nouns, representing 163 noun occurrences, were looked up in OpenCyc, resulting in 11,637 unique assertions added to the seven graphs.

After relevant background information is added to the graph, the syntax-semantics mapping rules are applied. For example, the rule nounPhraseToInstance is applicable to “a white van.” This rule states that, “If a common noun has a determiner as a dependent, and the text of the noun is txt, then the common noun token is an instance of the txt type." This rule adds the assertion, (Isa n155 van) and deletes the (determinant) assertion (det n155 n121). Later, the rule colorProperty applies. This rule states that, “A token with a noun (nn) or an adjective modifier (amod) that is a color has that color as its color property.” This rule adds the assertion, (color n155 white), and deletes (amod n155 n153). A later rule, removeTextWhenIsa, deletes (TextOf van n155), because n155 no longer denotes an occurrence of the word “van”, but an entity that is an instance of the category, van (see Figure 2). There are currently 23 syntax-semantics mapping rules that apply to the seven soft messages at least once. They apply a total of 687 times to the seven messages.

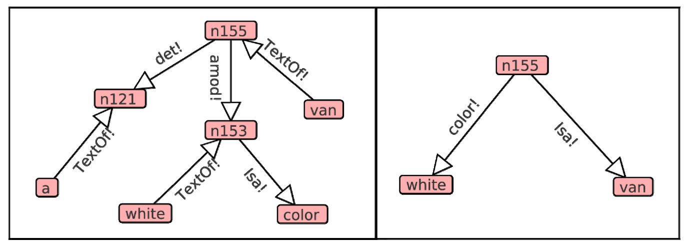

Figure 2. The Syntactic Graph for “a white van” on the Left is Translated to a Semantic Graph on the Right.

The final output of Tractor is a semantic graph representing the information contained in a single message. Currently, these graphs still contain syntactic information about the act and events, but the entities are described semantically, laying the ground for later stages of soft information processing.

V. PROPOSITIONAL GRAPH CONVERSION

Nodes of the semantic graphs generated by Tractor are linked together by syntactic information. The combination of syntactic information in addition to the semantic information increases the dimensionality of these semantic graphs. The downstream graph analytic algorithms of data association and graph matching have runtime that is a function of the number of nodes and edges in a graph. Hence this step was introduced to reduce the size of semantic graph and generate a smaller dimensional attributed graph while retaining the semantic information generated by the Tractor process.

The process of converting propositional graphs to attributed graphs is a rule-based approach that follows the syntactic structure of the graphs. The semantic graph for "a white van", shown in Figure 2 is converted into an attributed graph shown in Figure 3. The rule applied here takes two nodes (white and van) and converts them into a single node (van) with an attribute color with value white.

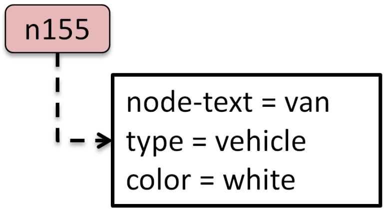

Figure 3. The Attributed Graph for "a white van".

VI. UNCERTAINTY ALIGNMENT

Human observation has the disadvantage of being more difficult to calibrate than its hard data counterpart. This is due to the context dependent and often qualitative nature of human observation. To overcome this difficulty an uncertainty alignment process was developed to apply the appropriate uncertainty representations for quantitative and qualitative observations [15]. An extensive literature review on human observational error characteristics yielded 3 categories of observational bias and variance influencing factors: observer characteristics, environment characteristics and observed value characteristics. Utilizing data provided in the identified articles, context-aware error models were developed which describe the observational bias, variance and distribution family of an attribute observation under the given observation conditions.

Incoming observations are run through the uncertainty alignment process which applies the appropriate bias and variance to the given observation (based on the identification of the appropriate error model). The developed distribution on the observed attributes value is then utilized in downstream processing to enable matching of inexact attribute values as well as for the preservation of uncertainty through the situation awareness process. Examples of attributes which are aligned through the uncertainty alignment process are BBS-2 and BBS-12 where the unknown man is qualitatively described as being of "medium" height. The upstream process properly identify "medium" as the height attribute value of this male which is then converted to the appropriate fuzzy height representation in the uncertainty alignment process.

VII. HARD DATA TRACK CONVERSION

Hard data is typically very precise and well characterized. We associate hard data with soft data by first abstracting away many of the details present in the hard data to obtain a form which is comparable to the soft processed data. This "soft-compatible" form is added to the processed soft-message stream and associated using a graph-based association process. An alternative method of modifying the soft messages to be comparable with the hard data was considered, however this approach requires imposing artificial precision to the soft data. Given this, it was decided to abstract detail away from hard data, rather than attempt to force hard data characteristics onto the soft data.

For geo-location data, this conversion is accomplished by first determining what named location corresponds to a given coordinate (hard datum). An attributed graph can then be built, an example of which is shown in Figure 4 where the center node represents some unknown tracked object(s), arcs indicate visited relationships and the outer nodes are locations.

Example geo-location log:
(1/1/2011) 33° 7'51.11"N, 41° 7'1.70"E
(1/2/2011) 33°31'51.86"N, 36°22'24.41"E
(1/3/2011) 31°20'8.36"N, 40°22'53.00"E
(1/4/2011) 33°18'14.32"N, 44°23'43.22"E

Corresponding soft-compatible data:
(1/1/2011) Visited: Iraq
(1/2/2011) Visited: Dimashq, Syria
(1/3/2011) Visited: Saudi Arabia
(1/4/2011) Visited: Baghdad, Iraq

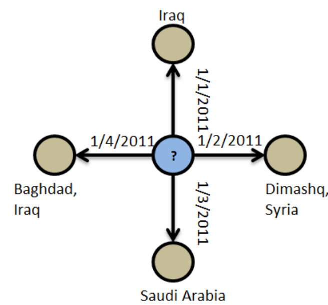

Figure 4. The Graphical Form for the Provided Coordinate Data.

VIII. DATA ASSOCIATION SCORING

Hard and soft uncertainty aligned attributed graphs form the input to the association scoring process. Each graph represents a single report within which co-references are assumed to have been resolved in the NLP processes. Since within message resolution is complete only cross message associations remain. A filtering process ensures only similar cross report entities, relationships and attributes are scored against one another (e.g., a person is not scored against a vehicle). This filtering process also indicates to the downstream association process which nodes and edges should not be merged based on type dissimilarity. Cross report similarity values are calculated at the attribute to attribute level and aggregated to form node-node and edge-edge similarity scores.

Text attributes are scored using a semantic similarity metric on the WordNet hierarchy which has been shown to provide a good approximation to a human's judgment of similarity [16]. Numerical attributes are scored using a fuzzy similarity metric which provides an uncertain similarity value indicating the possible range of similarity [17]. For the purpose of association, which requires crisp similarity values a fuzzy number ranking method is utilized in the case of an uncertain node-node or edge-edge score [18]. The similarity scores are then adjusted by a learned constant by uniformly subtracting/adding the constant value from/to all scores, and are then normalized to be in the range of -1 to 1 as required by the data association algorithm. The scores used here are not necessarily portable; different scenarios and applications will likely require retraining for the specific domain under examination.

IX. DATA ASSOCIATION

The problem of data association has been extensively studied with respect to physical sensors. An example is the problem of multi-target tracking, or the partitioning of observations into tracks [19][20]. This project utilizes graph association, which is an extension of the multidimensional assignment problem to accept relational data. Graph association is the central component of data association used here. To solve graph association is to merge many graphs that collectively describe a set of possibly repetitive entities and relationships into a single graph that contains unique entities and relationships (the cumulative data graph) by identifying pairs of nodes and edges that represent the same entities and relationships respectively.

The BBS demonstrates the benefit of using a graph association process, for example take the following two graphs that are produced from the soft messages BBS-2 and BBS-12 respectively. Underlying graph association is a set of scores that measure how likely two nodes or edges are to be referring to the same object or relationship respectively. Scores here are generated by the process described in Section VIII. In general, scores can range from negative to positive 1, reflecting the spectrum of two objects (relationships) being very unlikely to be the same entity (relationship) to almost certainly the same.

For instance, the man node in this example is scored -0.36 with the Mahmoud Azhour node even though Azhour is a man, reflecting the rarity at which a randomly selected man will in truth be Azhour. However, in this case, we also know that both the man and Azhour are related to a hair object, and through this relationship enough evidence exists for them to be associated. Graph association seeks to maximize the sum of scores for selected associations, so without the information added by the "part" relationship, man and Azhour would not be associated and subsequently merged since they had a negative score. Graph association would not merge these nodes since not merging them yields a benefit of zero by default, and 0 > -0.36.

Nodes:

	Hair	Man
Mahmoud Azhour	-0.76	-0.36
Jacket	-0.49	-0.16
Hair	0.10	-0.39

Edges:

	hasPart
With	0.80
Wearing	0.09

As with the multidimensional assignment problem, graph association is a difficult problem so a Lagrangian based heuristic for its solution is used. At completion of this heuristic, a merging step is used to combine the many graphs into a single graph by merging pairs of nodes and edges decided to represent same entities and relationships.

Structurally, this process is straightforward. The merging algorithm iterates over all pairs of elements found to be associated, replacing the pair by a single element and updating any connections in the graph to point to this new element.

The attributes placed on the merged element is the union of the attribute sets from the two elements that were merged. When there is disagreement between the elements' attribute sets (e.g., one node has a height attribute of tall being merged with another having height short) the conflict is resolved by merging the possibility distribution representation of their uncertainties following the methodology of [21]. For more information on graph association, please refer to [24].

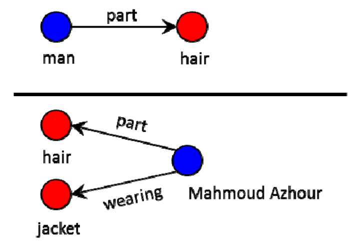

Figure 5. Cross-message Coreference Example.

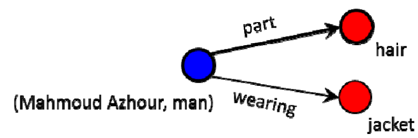

Figure 6. Merged Association Graph.

X. GRAPH MATCHING

The situation assessment process is a stochastic graph matching algorithm. The graph matching algorithm attempts to locate the most similar instance of a template graph or situation of interest within the cumulative data graph. These situations of interest can include priority intelligence requirements (PIR) provided to an intelligence analyst from a commander or standing threats which the analyst must be consistently vigilant of. The template may contain uncertain graph elements in the same format as the cumulative data graph. The graph matching process utilized here is a stochastic truncated search tree approach [22][22], using a similar scoring and ranking mechanism described in Section VIII. The outputs of the graph matching process are the similarity ordered matching results with an indication of the uncertainty in the overall and elemental similarities. The preservation of pedigree information allows the analyst to view matched graph elements' source directly in the original reports. This ability enables a more complete understanding of the matched result and draws the analyst's attention to other potentially significant data which was not contained in the original template.

Graph matching is performed incrementally on existing templates, avoiding the recalculation of similarity values and re-construction of the entire search tree [23]. Only search tree branchings which would have formed differently are rebuilt in the incremental graph matching algorithm, leading to a significant performance benefit. A situation of interest within the BBS is whether a bomber has rendezvoused with another person at a particular location (see Figure 7). This template is a standing threat which the analyst would like to be alerted of as matching cumulative data graph information becomes available. As the messages flow through the processing stream the incremental graph matching algorithm is executed and graph matching results updated. After BBS-7 the top similarity score is ranked 0.75. However, after BBS-8 is added to the cumulative data graph, placing the green SUV at the apartment building a perfect match of 1.0 is attained.

The recognition of the rendezvous of known bomber Sayed Azhour with his brother Mahmoud at the Harma Apartments allow actions to be taken to successfully bust the bomber. This match exemplifies the importance of the joint consideration of hard and soft data. The LIDAR+PTZ camera track identifies Sayed's green SUV at the Harma Apartments while the police tip-line report of the apartment resident identifies the white SUV as being present at the Harma Apartments. The successful associations of the Harma Apartment location (BBS-5 and BBS-8) and Mahmoud Azhour with the white SUV (BBS-2 and BBS-5) enable this match.

Figure 7. Template Graph Representing Bomber Rendezvous.

XI. CONCLUSIONS AND FUTURE WORK

This paper has presented a unified framework for the fusion of hard and soft information. Hard and soft data is successfully integrated further upstream than in our previous framework [5], allowing more elemental (at the entity and relationship association level) connections to be made across the data modalities. The example BBS illustrates the synergistic effect of hard and soft integration, providing better situation awareness than would have been attainable without cross-modality fusion.

Future work includes the continuation of the conversion of syntactic information to semantic information, better description and attribution of actions and events, validation and expansion of the context-aware human error models, the development of an incremental data association methodology and increasing the flexibility and efficiency of the graph analytic techniques.

ACKNOWLEDGMENT

This work has been supported by a Multidisciplinary University Research Initiative (MURI) grant (Number W911NF-09-1-0392) for "Unified Research on Network-based Hard/Soft Information Fusion," issued by the US Army Research Office (ARO) under the program management of Dr. John Lavery. We gratefully appreciate this support.

REFERENCES

[1] "Unified Research on Network-based Hard/Soft Information Fusion", Multidisciplinary University Research Initiative (MURI) grant (Number W911NF-09-1-0392) by the US Army Research Office (ARO) to University at Buffalo (SUNY) and partner institutions.

[2] United States. Department of the Army. FM 34-130 Intelligence Preparation of the Battlefield. 1994.

[3] United States. Department of the Army. *FM 101-5-1 Operational Terms and Graphics*. 1997.

[4] D.L. Hall., and J. Llinas. "An Introduction to Multisensor Data Fusion." *Proceedings of the IEEE* 85.1 (1997): 6-23.

[5] J. Llinas, R. Nagi, D. Hall, and J. Lavery. "A multi-disciplinary university research initiative in hard and soft information fusion: overview, research strategies and initial results." Proc. of 13th Conference on Information Fusion (FUSION), Edinburgh.

[6] M. Prentice, M. Kandefer, and S. C. Shapiro, "Tractor: A framework for soft information fusion," *Proceedings of the 13th International Conference on Information Fusion (Fusion 2010),* 2010, Th3.2.2

[7] M. Prentice and S. C. Shapiro, "Using propositional graphs for soft information fusion," *Proceedings of the 14th International Conference on Information Fusion (Fusion 2011),* 2011, 522-528.

[8] S. C. Shapiro and W. J. Rapaport, "The SNePS family," *Computers & Mathematics with Applications 23*, 2-5 (January-March,1992), 243-275. Reprinted in F. Lehmann, Ed. *Semantic Networks in Artificial Intelligence*, Pergamon Press, Oxford, 1992, 243-275.

[9] D. R. Schlegel and S. C. Shapiro, "Visually interacting with a knowledge base using frames, logic, and propositional graphs." In M. Croitoru, J. Howse,S. Rudolph, and N. Wilson, Eds., Graph Structures for Knowledge Representation and Reasoning, Lecture Notes in Artificial Intelligence 7205, Springer-Verlag, Berlin, in press.

[10] H. Cunningham, et al. Text Processing with GATE (Version 6). University of Sheffield Department of Computer Science. 15 April 2011. ISBN 0956599311.

[11] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan. "GATE: A framework and graphical development environment for robust NLP tools and applications." Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia, July 2002.

[12] Snowball, http://snowball.tartarus.org/.

[13] M.-C. de Marneffe and C. D. Manning, "The Stanford typed dependencies representation," COLING 2008 Workshop on Cross-framework and Cross-domain Parser Evaluation, 2008.

[14] OpenCyc.org, http://www.opencyc.org/.

[15] M. P. Jenkins, G. A. Gross., A. M. Bisantz, and R. Nagi, (2011). "Towards context-aware hard/soft information fusion: Incorporation of situationally qualified human observations into a fusion process for intelligence analysis." In *Proceedings of the 2011 IEEE First International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, Feb. 22-24, Miami Beach, FL, pp. 74-81.

[16] J. J. Jiang, and D. W. Conrath. "Semantic similarity based on corpus statistics and lexical taxonomy." Proc. of International Conference Research on Computational Linguistics, Taiwan. 1997.

[17] D. Guha and D. Chakraborty, "A new approach to fuzzy distance measure and similarity measure between two generalized fuzzy numbers," *Applied Soft Computing,* vol. 10, pp. 90-99, 2010.

[18] S.-M. Chen and J.-H. Chen, "Fuzzy risk analysis based on ranking generalized fuzzy numbers with different heights and different spreads," *Expert Systems with Applications,* vol. 36, pp. 6833-6842, 2009.

[19] A. Poore, S. Lu, and B. Suchomel. "Data association using multiple-frame assignments." In Martin Liggins, David Hall, and James Llinas, editors, Handbook of Multisensor Data Fusion, chapter 13, pages 299-318. CRC Press, Taylor & Francis Group, 2 edition, 2009.

[20] A. Poore and N. Rijavec. "A lagrangian relaxation algorithm for multidimensional assignment problems arising from multitarget tracking." SIAM Journal of Optimization, 3(3):544-563, 1993.

[21] R. Yager (Machine Intelligence Institute, Iona College). Conditional Approach to Possibility-Probability Fusion. 2011. 30p Report No.: MII-3021R.

[22] G. Gross, R. Nagi, and K. Sambhoos, "Soft information, dirty graphs and uncertainty representation/processing for situation understanding," in The 13th International Conference on Information Fusion. 2010: Edinburgh, Scotland.

[23] G. Gross, R. Nagi, and K. Sambhoos, "Continuous preservation of situation awareness through incremental/stochastic graphical methods," in The 14th International Conference on Information Fusion. 2011: Chicago, IL.

[24] G.Tauer, R. Nagi, and M. Sudit. The Graph Association Problem: Mathematical Models and a Lagrangian Heuristic. Under Review, submitted to Navel Research Logistics. 2011.

APPENDIX A – BOMBER BUSTER SCENARIO MESSAGES

Message ID	Type	Time	Source	Content
BBS-1	Soft	01/31/2010 0700 hrs	Local newspaper	Al Sabah newspaper reports that in response to the new government policy, local presidential candidate Azam Al-Azhar has called for a protest at the Second District Courthouse. Al-Azhar said he would personally attend this protest, and that local residents should expect to see his black SUV arrive at the Courthouse at around 1800 hrs.
BBS-2	Soft	01/31/2010 1720 hrs	Police tip-line	A Baghdad van rental company owner reports that he just rented a van to a medium height man who was wearing a tan jacket. The owner's wife recognized the man as Mahmoud Azhour, who she believes is the brother of known bomber Sayed Azhour. The rented vehicle is a white van with license plate number 72751.
BBS-3	Soft	01/31/2010 1745 hrs	Intercepted cell call	Blue Team intercepted a cell phone call from Mohammed Wali to Sayed Azhour, saying that he is driving a red car, and is currently parked by the Second District Courthouse. Sayed Azhour said that he was driving a green SUV.
BBS-4	Hard	01/31/2010 1800 hrs	Camera by Second District Courthouse	Arrival to Government Building of Black SUV with license plate number 4321.*
BBS-5	Soft	01/31/2010 1802 hrs	Police tip-line	A resident of Hamra Apartments reports that a white van is blocking his parking space in front of his apartment building. The resident did not write down the license plate number, but noticed that the white van had a sticker on it indicating that the van was a rental.
BBS-6	Soft	01/31/2010 1803 hrs	Blue Team	Blue Team reports they are in position at the Second District Courthouse. They have noticed both a black SUV and a red car parked near the courthouse.
BBS-7	Hard	01/31/2010 1803 hrs	Camera by Second District Courthouse	Arrival to Government Building of Blue SUV.*
BBS-8	Hard	01/31/2010 1805 hrs	LIDAR+PTZ Camera fusion	Vehicle tracked by LIDAR to apartment building. PTZ Camera IDs the track as a green SUV.*
BBS-9	Soft	01/31/2010 1808 hrs	Blue Team	Blue team has confirmed that there is a large protest at the Second District Courthouse.
BBS-10	Hard	01/31/2010 1808 hrs	Acoustic sensor	Acoustic sensor detects car doors opening and shutting.*
BBS-11	Hard	01/31/2010 1815 hrs	Camera by Second District Courthouse	Arrival to Government Building of White Van with license plate 72751.*
BBS-12	Soft	01/31/2010 1817 hrs	Blue Team	Blue team reports a medium height man with dark hair wearing a tan jacket just entered a red car by the Second District Courthouse. They are not sure where the man came from.

* - hard data content is presented as a linguistic summary of the fused sensor data