A First Step Toward the Formal Analysis of Solution Quality in FA/C Distributed Interpretation Systems

Norman Carver and Victor Lesser

Department of Computer Science, LGRC-243
University of Massachusetts
Amherst, Massachusetts 01003
(carver@cs.umass.edu, lesser@cs.umass.edu)

Abstract

In the functionally-accurate, cooperative (FA/C) distributed problem-solving paradigm, agents produce tentative, partial results based on local information only, and then exploit the constraints among these local results to resolve uncertainties and global inconsistencies. However, there has never been any formal analysis of the quality of the solutions that are produced by the approach or of the conditions that are necessary for the approach to be successful. This paper represents a first step in formally analyzing the quality of solutions that can be produced by FA/C systems, within the context of distributed interpretation. Two theorems that compare the quality of solutions produced by a distributed system to those produced by an equivalent centralized system are presented. The theorems relate solution quality to agent problem-solving and coordination strategies. The analysis is based on an abstract model of the DRESUN system for distributed sensor interpretation. While the paper concentrates on sensor interpretation, we expect to extend the work to apply to FA/C systems in general.

This work was supported by the Department of the Navy, Office of the Chief of Naval Research, under contract N00014-92-J-1450. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

1 Introduction

In the *functionally accurate, cooperative* (FA/C) systems paradigm for cooperative distributed problem solving (CDPS) [6, 8], agents produce tentative, partial results based on incomplete local information. These partial results are then exchanged among the agents, which exploit the constraints that exist among the results to resolve local uncertainties and global inconsistencies. While several systems that use the FA/C approach have been built (e.g., [2, 7]), there has never been any formal analysis of the quality of the solutions that are produced by the approach or of the conditions that are necessary for the approach to be successful.

This paper represents a first step in formally analyzing the quality of solutions that can be produced by an FA/C system. Two theorems are presented that compare the quality of solutions produced by a distributed system to the solutions that would be produced by an equivalent centralized system. The theorems relate solution quality to agent problem-solving and coordination strategies. They show that there are conditions under which it is possible to guarantee that the distributed system produces a solution that is comparable to the centralized solution, and other conditions under which there is merely some probability of obtaining such a solution. The analysis assumes the capabilities of an abstract model of the DRESUN system for distributed sensor interpretation [2, 3]. While the paper concentrates on distributed sensor interpretation, we expect to extend the work to be able to make general statements about the FA/C approach. The paper is also limited to the issue of solution quality at termination; we are pursuing both empirical and analytic approaches to address other issues in FA/C problem solving.[1]

This work is different from most DAI work dealing with global consistency of agent beliefs (e.g., [4, 5]), which has focused on methods for automatically maintaining (some level of) consistency. Here, we are interested in analyzing the consequences of not maintaining consistency (even within individual agents) in terms of its effect on solution quality. This is an important issue since for problems like distributed sensor interpretation, communication and computation considerations often make it impractical to maintain complete concistency.

In the next section we describe our model of sensor interpretation and discuss the issues that arise in determining the most likely interpretation. Section 3 examines the distributed problem-solving model that we assume for our analysis. Section 4 then introduces the notation that will be used in the theorems and proofs of Section 5. The paper concludes with a summary of our conclusions and future research plans.

[1] One of the key issues in the design of FA/C systems is the role that agent architectures and coordination strategies play in allowing a wide range of inconsistencies to be resolved without requiring excessive communication of data and hypotheses among the agents [2, 3].

2 Distributed Sensor Interpretation

By sensor interpretation, we mean the determination of high-level, conceptual explanations of sensor data. Our model of the interpretation process will be essentially that used in [1]: interpretation hypotheses are incrementally constructed via *abductive inferences*, based on a causal model that defines the relationships among the data types and abstraction types. For each type T, the causal model defines the type's support, $\mathcal{S}_T$, and its possible explanations, $\mathcal{E}_T$ (if T is a data type then $\mathcal{S}_T$ is empty, likewise if T is a top-level type then $\mathcal{E}_T$ is empty). $\mathcal{S}_T = \{S_i\}$, a set of *type instance specifications*—i.e., each S_i is an interpretation type plus a set of constraints that specify the consistent attribute values for hypotheses of that type and hypotheses of type T. $\mathcal{E}_T = \{E_i\}$, a set of types, each of which might explain some type T hypothesis.

An interpretation system makes abductive inferences that identify possible explanations for a piece of data or a hypothesis. Thus, every hypothesis h with $type(h) = T$ is the result of a set of abductive inferences each of the form $h_i \Rightarrow h$, where $type(h_i) \in \mathcal{S}_T$. Conversely, each hypothesis h with $type(h) = T$ may be part of a set of abductive inferences of the form $h \Rightarrow h_j$, where $type(h_j) \in \mathcal{E}_T$. We will use the following notation: *support-of*$(h) = \{h_i\}$, the (immediate) support for h; $supports(h, h_i)$; and $explains(h, h_j)$. In addition to its immediate support, each hypothesis must eventually be supported by sensor data via a chain of abductive inferences $d \Rightarrow h_1 \Rightarrow \ldots \Rightarrow h_m$, where d is data and $supports(h_1, d)$, $supports(h_l, h_{l-1})$. To deal with these inference chains, we will use the notation *support-of**(h) to refer to that data that (eventually) supports h, as well as the analogous *supports**(h_m, d), and *explains**(d, h_m).

Abductive inferences are uncertain rather than logically correct inferences that provide *evidence* for the hypotheses rather than conclusively proving them. The key source of uncertainty for any hypothesis is the possibility of alternative explanations for the data that supports the hypothesis. Even if *complete* support can be found for a hypothesis (i.e., there is a hypothesis or datum in *support-of*(h) that corresponds to each $S_i \in \mathcal{S}_T$), the hypothesis may still be uncertain as a result of competing, alternative explanations for *support-of**(h). Furthermore, because hypotheses are incrementally constructed, they may not have complete support at any point in time, even if complete support could be found in the available data (complete support might not be able to be found even if the hypothesis is correct due to masking, environmental disturbances, or sensor errors).

A *solution* to an interpretation problem is an explanation of what caused the available data. In general, this will be a *composite* interpretation, composed of a set of hypotheses whose types are from a specified subset of the interpretation types (the *explanation corpus* [9]), each of which explains some subset of the data. In a centralized interpretation system, all of the data is available to the single agent. In a distributed system, each agent has (direct) access to data from only a subset of the sensors, and each sensor is associated with a single agent. As a result, each agent monitors only a portion of the overall area of interest, and agents' local solutions must be combined in order to construct a global solution. Construction of a global solution may not be straightforward, however, because the local solutions are often not independent and may in fact be inconsistent because they are based on different incomplete subsets of the data. Agent solutions are interdependent whenever data (evidence) for a hypothesis is spread among multiple agents or when agent areas of interest overlap as a result of overlapping sensor coverage.

In this paper we are interested in analyzing the quality of the global solutions that can be produced by FA/C distributed interpretation systems. We will analyze solution quality relative to what we term the "globally best solution." Ideally, this should be the *most probable explanation* (MPE) [9] given all of the available data.[2] The problem with defining the globally best solution to be the global MPE is that for many interpretation problems it is impractical to compute the MPE. Because our main interest is the effect of FA/C distributed problem solving, we have chosen to compare distributed system solutions to those of an *equivalent* centralized system. By this we will mean the solution that would be produced by a centralized (single-agent) system that has access to all of the data of the distributed system and uses the same local problem-solving strategies.

The complexity of computing the MPE can be understood by considering the differences between interpretation problems and the kinds of problems that are typically studied in research on abductive inference and probabilistic network inference (e.g., [9, 10]). For simplicity, we will refer to these problems as *diagnosis problems.* The key difference is that these problems are propositional while interpretation is not, in general. In other words, diagnosis problems have a fixed set of causes and possible findings, with fixed relations among them. By contrast, while interpretation problems have a fixed set of top-level cause *types* and a fixed set of data *types*, they can have an indeterminate number of instances of any of the types. In particular, there can be an indeterminate number of instances of any top-level cause (e.g., vehicles in a vehicle monitoring system). This leads to the problem of *correlation ambiguity* (it is ambiguous/uncertain which potential explanation hypothesis a support instance should be associated with), which results in a combinatorial explosion of possible explanations for a data set. Furthermore, interpretation systems are often faced with large volumes of data.

Because these factors make it impractical to determine the MPE, interpretation systems must use heuristic, satisficing approaches to construct solutions. Thus, interpretation systems usually do

[2] Note that the MPE is not necessarily the *correct* interpretation—i.e., what actually produced the data—it is simply the most likely interpretation. The MPE can be "incorrect" when the characteristics of the available data are unusual or when the system has poor models of the domain.

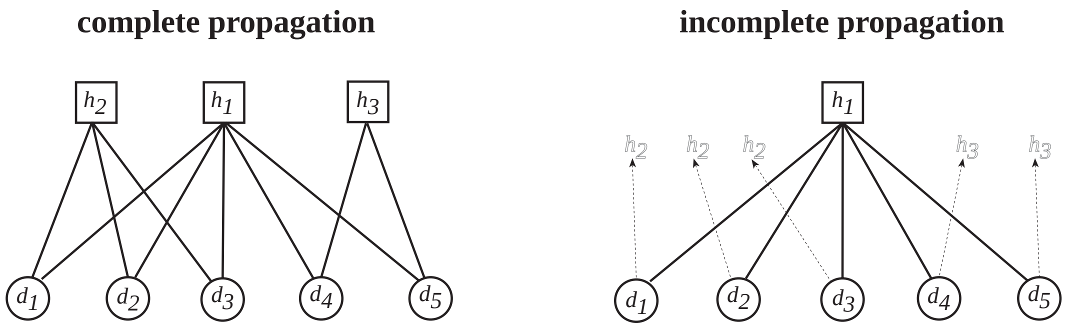

Figure 1: An example of incomplete evidence propagation.
In the complete propagation case, the system not only has created the most probable explanation, h_1, it also has created the alternative explanations, h_2 and h_3 (using the most complete support possible). This allows the system to determine the conditional probability of h_1 given the available data $\{d_1, \ldots, d_5\}$. In the incomplete propagation case, the alternative explanations for h_1 have not been created. This means that the belief computed for h_1 is only an approximation of the true conditional probability since the likelihood of the alternative explanations has not been correctly considered (h_1 is still uncertain, though, because the possibility of alternative type 2 and 3 explanations for each piece of supporting data is known, as are the a priori likelihoods of these explanations). (Note that for simplicitly, the figure does not show the numerous incompletely supported versions of the hypotheses that also would have been created in the case of complete propagation.)

not process every piece of data available to them nor construct all the possible explanations for a data set, and they may assemble solutions from hypotheses whose belief ratings surpass some *acceptance threshold* (rather than verifying that they are the most likely). These kinds of approaches mean that solutions are only approximations of the MPE. For example, if a system does not create every possible interpretation of its data (i.e., every possible hypothesis h such that *explains**(D, h)), this not only limits the interpretations that can be considered, it also results in hypothesis belief ratings being only approximations of the conditional probabilities of the hypotheses. This is because incomplete hypothesis construction is equivalent to incomplete propagation/evaluation of evidence—see Figure 1.

Despite their drawbacks, such approaches can be effective because: systems are typically interested in only certain types of phenomena out of all the environmental phenomena for which there are models (i.e., answer versus nonanswer types); data may be redundant due to the existence of multiple sensors; and it may not be necessary to process every piece of relevant data and make every evidential inference in order to be sufficiently certain of interpretations. Of course, since the use of heuristic strategies affects solution quality, these strategies must be taken into account in our analyses. For example, it is the acceptance threshold, along with other factors like the phenomena of interest, that are specified by the *termination criteria*—the conditions that must be met for termination of problem solving. The RESUN/DRESUN architecture provides great flexibility for implementing interpretation strategies [1, 2].

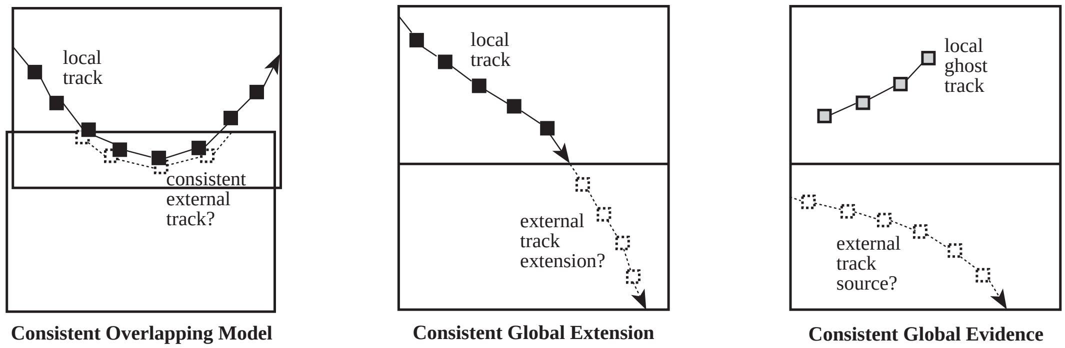

Figure 2: Examples of global consistency SOUs for vehicle monitoring.

3 The DRESUN Model

In an FA/C system, there must be some mechanism to drive interactions among the agents so that incorrect and inconsistent local solutions can be detected and dealt with. Ideally, this would be accomplished with a mechanism that allowed agents to understand where there are constraints among their subproblems, so that information interchange could be highly directed. DRESUN [1, 2, 3] provides this capability, and it will form the basis for our model of the capabilities of an FA/C agent. DRESUN agents create symbolic source of uncertainty statements (SOUs) whenever it is determined that a local hypothesis can obtain evidence from another agent—i.e., whenever a subproblem interaction (constraint) is detected. Global consistency SOUs are viewed as sources of uncertainty about the correctness of an agent's local solution because they represent unresolved questions about the global consistency of the solution.

To see that it is possible to identify all possible subproblem interactions, consider that there are just three classes of global interactions in sensor interpretation problems: interpretations in regions of overlapping interest among agents must be consistent, "continuous" hypotheses that would extend into other agents' areas must have consistent external extensions, and hypotheses that require evidence that could be in another agent's area must have consistent external evidence. Instances of these situations can be detected given the domain model and knowledge of the organization of agent interest areas. DRESUN uses three global consistency SOUs to denote instances of these global interactions: *consistent-overlapping-model*, *consistent-global-extension*, and *consistent-global-evidence*. Figure 2 shows examples of situations involving each of these SOUs.

While DRESUN agents have a representation of all inter-agent interactions (for the set of locally created hypotheses), we must now consider what should be done with this information—i.e., what are appropriate coordination strategies for *resolving* the global SOUs? Resolution of a global SOU involves exchanging information among the associated agents so as to effectively propagate evidence between their hypothesis (belief) networks. An example of the resolution of a global SOU is shown

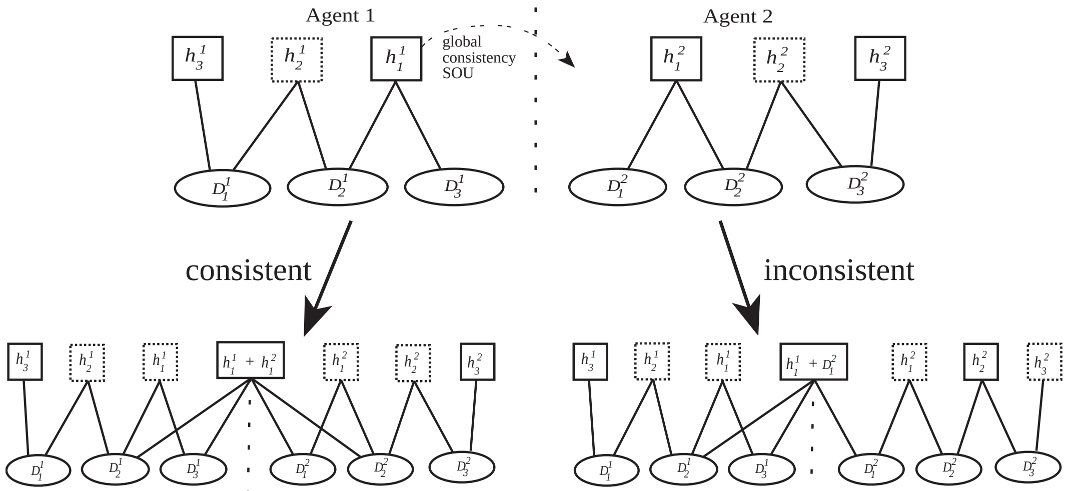

Figure 3: An example of the resolution of a global consistency SOU.
When there is a consistent explanation in the external agent, resolution of the global SOU associated with h_1^1 results in the creation of a merged hypothesis as a new alternative explanation in each agent. When the local hypothesis is inconsistent with hypotheses in the external agent, new alternatives may be created (as shown here). When the local hypothesis is inconsistent with the data in the external agent, new evidential links are created to represent the contradictory evidence.

in Figure 3. Resolution of global SOUs is analogous to (intra-agent) evidence propagation, and as with evidence propagation there are a range of strategies that may be used to determine which global SOUs to pursue and how completely to propagate their effects.[3] The most comprehensive strategy is for all global SOUs to be *completely* resolved (complete resolution means that all possible evidential inferences and hypotheses that could result from the interaction are created). A less comprehensive strategy (that still insures the ability to merge the local solutions) is to resolve only those SOUs that are associated with each agent's best local solution and to pursue propagation using only the best solution of the external agents (of course, if the agents' solutions are not consistent, this could trigger additional inferences that would create and explore newly more likely interpretations).

4 Notation

For the centralized (single-agent) case, we will use the following notation:

- $\mathcal{D}$ is the complete set of data available to the system, and D_i denotes some subset of $\mathcal{D}$ ($D_i \subseteq \mathcal{D}$).
- C_i denotes an interpretation *context*. Each context refers to a specific data subset that has been processed—i.e., each C_i is associated with a $D_i \subseteq \mathcal{D}$.
- $BEL(h, C_i)$ is the *belief* in hypothesis h given context C_i. In other words, it is $P(h \mid D_i)$, where D_i is the data subset associated with context C_i. In general, $BEL(h, D_i) \neq BEL(h, \mathcal{D})$ when $D_i \subset \mathcal{D}$. Further, because interpretation domains are not typically monotone, $BEL(h, D_i)$ may be less than or greater than $BEL(h, \mathcal{D})$.

[3]Different coordination strategies for how and *when* to communicate can greatly affect the efficiency of the CDPS process [3]. However, for this paper, we are interested only in solution quality.

- $\widehat{BEL}(h, C_i)$ is the approximate value of $BEL(h, C_i)$ that is computed given the evidence propagation strategy being used—i.e., the hypothesis' belief rating. $\widehat{BEL}(h, C_i) \neq BEL(h, C_i)$ because in general the control component will not pursue all possible evidential inferences from the examined data since this can lead to the construction of a combinatorial number of alternative hypotheses. As above, $\widehat{BEL}(h, C_i)$ will be written as $\widehat{BEL}(h)$ when the context is clear.
- $I(C_i)$ is the set of all possible (composite) interpretations of the data subset associated with context C_i. Each $I_j(C_i) \in I(C_i)$ is a set of hypotheses that can explain all of the data associated with C_i. We will use $I(\mathcal{D})$ to represent the set of all possible interpretations of a data set.
- $\mathcal{I}(C_i)$ is the true MPE of the data subset associated with context C_i.
- $\hat{\mathcal{I}}(C_i)$ is the estimated best explanation of the data subset associated with context C_i—given the evidence propagation and satisficing solution selection strategy being used.

For the distributed, multi-agent case, modifications must be made to account for the distribution of data and hypotheses among multiple agents. The following notation will be used for the distributed case:

- $\mathcal{A}$ is the set of agents $\{A_1, A_2, \ldots\}$, with their interest area specifications.
- $\mathcal{D}^i$ is the complete set of data available (directly) to only agent A_i—i.e., the complete set of data that is available from agent A_i's own sensors. D^i_j denotes some subset, j, of this data.
- $\mathcal{D}^G$ refers to the complete set of globally available data—i.e., the combined data from all of the agents, $\bigcup_{A_i \in \mathcal{A}} \mathcal{D}^i$. $\mathcal{D}^G = \mathcal{D}$ for an equivalent centralized system.
- C^i_j denotes an interpretation context for agent A_i. Each context denotes both some subset of the local data that has been processed by the agent, D^i_j, as well as any external data communciated from other agents.
- C^G_i will be used to refer to the global context—i.e., $C^G_j \equiv \bigcup_{A_i \in \mathcal{A}} C^i_j$.
- $BEL^i(h, C_j)$ is agent A_i's belief in hypothesis h given the context C^i_j.
- $\widehat{BEL}^i(h, C_j)$ is agent A_i's approximate value of $BEL^i(h, C_j)$ that is computed given the partial evidence propagation strategy being used by A_i (see the centralized section above for a discussion of the relation between $\widehat{BEL}()$ and $BEL()$).
- $\mathcal{I}^i(C_j)$ is the MPE for the data subset associated with context C^i_j.
- $\hat{\mathcal{I}}^i(C_j)$ is agent A_i's estimated best explanation of the data subset associated with context C^i_j—given the evidence propagation and satisficing solution selection strategy being used.
- $\hat{\mathcal{I}}^G(C_j)$ denotes the combined, global estimated best explanation of the data subset associated with the global context C^G_j. That is, $\hat{\mathcal{I}}^G(C_j) = \bigcup_{A_i \in \mathcal{A}} \hat{\mathcal{I}}^i(C_j)$. $\hat{\mathcal{I}}^G(C_j)$ is defined only when all the $\hat{\mathcal{I}}^i(C_j)$ are consistent, as in a final context, which satisfies the consistency termination criteria.

5 Theorems

To rephrase the main question we are interested in addressing: what is the relationship between $\hat{\mathcal{I}}^G(C^G_{f_d})$ and $\hat{\mathcal{I}}(C_{f_c})$, where $C^G_{f_d}$ and C_{f_c} represent possible final contexts (contexts that satisfy the termination criteria) for the distributed and centralized cases, respectively? For these theorems, the termination criteria will be phrased in terms of whether an interpretation is *acceptable* or not. We define the predicate $acceptable(\mathcal{I}(C))$ to mean that $\forall h \in \mathcal{I}(C), \widehat{BEL}(h) \geq AT$, where AT is the *acceptance threshold* $(0 < AT < 1)$.

5.1 Theorem 1

In this section we will examine the nature of global solutions assuming that all the data is processed and there is complete intra-agent evidence propagation. By this we mean that for all of the data that is available to an agent, the agent constructs all possible interpretation hypotheses and computes the (true) conditional probability of those hypotheses. For the centralized case, this means that the final context C_{f_c} refers to the complete data set $\mathcal{D}$, and that $\hat{\mathcal{I}}(C_{f_c}) = \mathcal{I}(C_{f_c})$—i.e., the final solution is the true MPE of the data. For the distributed case, the termination criteria must also specify the strategy for resolving the global SOUs. Here, we will specify that each agent must have resolved those global SOUs that are associated with its best solution, $\hat{\mathcal{I}}^G(C^G_d)$, and that if this solution is not acceptable, it must continue to resolve global SOUs until the solution is acceptable or all the global SOUs have been resolved.

Theorem 1: *Given complete intra-agent evidence propagation and the global propagation strategy specified above, $\forall \mathcal{D}$, if $acceptable(\hat{\mathcal{I}}(C_{f_c}))$ then $acceptable(\hat{\mathcal{I}}^G(C^G_{f_d}))$, where C_{f_c} and $C^G_{f_d}$ are final contexts that meet the specified centralized and distributed termination criteria for the $\mathcal{D}$, respectively.*[4] In other words, if a centralized system is able to produce an acceptable solution for a data set then so should a distributed DRESUN system using the specified strategy.

Proof: Assume that the theorem is false. This means that even though $acceptable(\hat{\mathcal{I}}(C_{f_c}))$, $\exists h \in \hat{\mathcal{I}}^G(C^G_{f_d}))$ such that $\widehat{BEL}^G(h, C^G_{f_d}) < AT$. Now, because $acceptable(\hat{\mathcal{I}}(C_{f_c}))$, $\exists \{h_j\} \subset \hat{\mathcal{I}}(C_{f_c})$ such that *explains*(support-of*(h)*, $\{h_j\})$ and $\forall h_j \in \{h_j\}, BEL(h_j) \geq AT$. In other words, the centralized system has one or more hypotheses that are part of its solution, that have acceptable belief ratings, and that explain the data underlying the unacceptable hypothesis h in the global interpretation. Since these hypotheses are alternative explanations for *support-of**(h), each of these

[4] Note that we can refer to the global interpretation $\hat{\mathcal{I}}^G(C^G_{f_d})$ rather than just individual agent interpretations $\hat{\mathcal{I}}^i(C^G_{f_d})$ here since in the final context $C^G_{f_d}$, individual agent solutions must be consistent (mergeable).

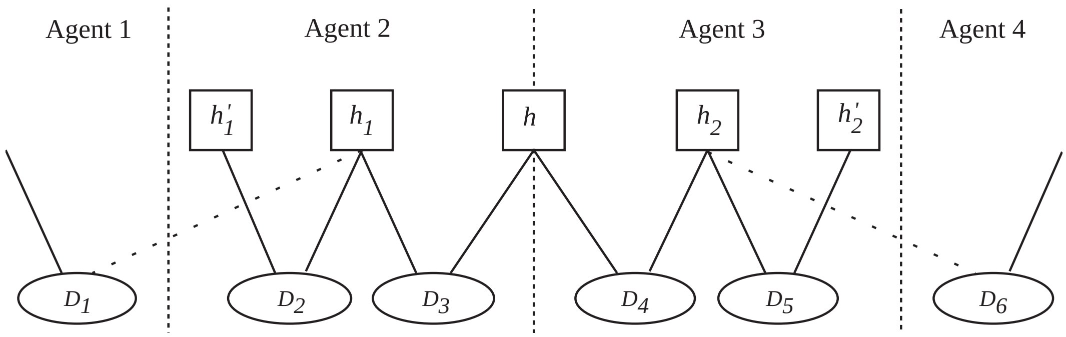

Figure 4: An example of the effect of partial resolution of the global SOUs.
Here, resolution of a global SOU btween agents 2 and 3 produced the hypothesis h, which resulted in $\{h, h_1', h_2'\}$ being judged as an acceptable explanation for agents 2 and 3. Because of this, the global SOUs associated with h_1 and h_2 were not resolved. However, resolution of these global SOUs could result in corroborating evidence being obtained from agents 1 and 4, which might make $\{h_1, h_2\}$ the best explanation.

hypotheses would also have been created by at least one of the agents that produced h.[5] So, for each h_j there is a corresponding hypothesis h_j^k that would have been created by A_k. Now, why is $\widehat{BEL}^k(h_j^k) < AT$ when $BEL(h_j) \geq AT$? There are three possibilities: (1) there exists data in one or more agents other than A_k that would corroborate h_j^k, but this data was not propagated to h_j^k;[6] (2) there exists data in one or more other agents that would contradict one or more alternatives to h_j^k; (3) both of the above may be true. Consider case 1: If there exists data in other agents that could support h_j^k, then A_k would have had global SOUs associated with h_j^k that identified the possible existence of this data (since global SOUs identify all interrelationships among the agents given the constructed hypotheses), but these SOUs were not resolved. This leads to a *contradiction* with our specified coordination strategy, however, since these SOUs should have been pursued when no acceptable explanation was found for the data common to h and h_j^k (remember that $h \in \hat{\mathcal{I}}^G(C_{f_d}^G)$ but $\widehat{BEL}^G(h, C_{f_d}^G) < AT$). Cases 2 and 3 lead to similar contradictions.

Thus, given complete intra-agent evidence propagation and the specified strategy for pursuing the global SOUs, if a centralized interpretation system can find an interpretation above a particular threshold then so can a DRESUN-based distributed interpretation system. This is a somewhat weak result, but it serves to establish a basic capability of the DRESUN architecture. Other FA/C

[5] Actually, the exact same hypothesis may not have been produced by a single agent since each of the agents may have only a portion of the data necessary to produce the version of the hypothesis with the most *complete* support possible. This issue arises because interpretation problems are not propositional, so as additional evidence is generated for a hypothesis the evidence both modifies the hypothesis' belief and refines the values of its attributes—creating new versions of the hypothesis (what [1] terms *extensions*). Thus, what we really mean is that a precursor, partial version of each alternative hypothesis would have been created by one or more of the agents. This is true because interpretation hypotheses can be created based on incomplete support, and our specified strategy was for each agent to create all the possible interpretations that are (even partially) supported by its own local data.

[6] The data cannot be in A_k since we have assumed complete intra-agent propagation.

distributed problem-solving architectures have not had this property because they have lacked a representation of the inter-agent constraints.

It is important to note that while we have shown that a distributed DRESUN system would find an "acceptable" solution, this interpretation may not be identical to that produced by the centralized system and may not be the MPE of the global data set. This is because an acceptable solution might be found without all the global SOUs having been resolved, which means that there would have been incomplete evidence propagation among the agents. As a result, hypothesis belief ratings would be only approximations of the true conditional probabilities and this could lead to alternative explanations being (incorrectly) rated more highly—see Figure 4. This is the reason that the proof used $BEL(h)$ for the centralized system, but $\widehat{BEL}^G(h)$ for the distributed system. The only way to guarantee that a distributed system produces a solution that is *identical* to that of the centralized system (and is the MPE of the globally available data) is to require that all of the global SOUs associated with every hypothesis in each agent be completely resolved.

5.2 Theorem 2

In this section we will begin to examine the nature of global solutions when complete intra-agent evidence propagation is impractical—a common situation for complex interpretation systems. This means that agents do not necessarily construct all possible interpretation hypotheses and so may compute only approximate belief ratings (conditional probabilities) for the hypotheses. Analyzing the solution quality given incomplete evidence propagation is complicated by the fact that a system's solutions can depend on the details of the strategies being used. We are not yet at the point where we can formally define and analyze the effects of different strategies. Thus, we will not specify a particular strategy for controlling evidence propagation here. Instead, we will simply assume that the same strategy is used in the single agent of the centralized case and in all of the agents of the distributed case. We will also assume the same coordination strategy (for resolving global SOUs) that was used in theorem 1, and we will continue to assume that all of the data directly available to each agent is (at least partially) processed by that agent.

Theorem 2: *Given incomplete intra-agent evidence propagation and the specified global propagation strategy, it is not the case that* $\forall\mathcal{D}$, *if* $acceptable(\hat{\mathcal{I}}(C_{f_c}))$ *then* $acceptable(\hat{\mathcal{I}}^G(C^G_{f_d}))$, *where* C_{f_c} *and* $C^G_{f_d}$ *are final contexts that meet the specified centralized and distributed termination criteria for the data set* $\mathcal{D}$, *respectively.* In other words, even if a centralized system is able to produce an acceptable solution for a data set, a distributed DRESUN system may not (even when the same basic propagation strategy is being used).

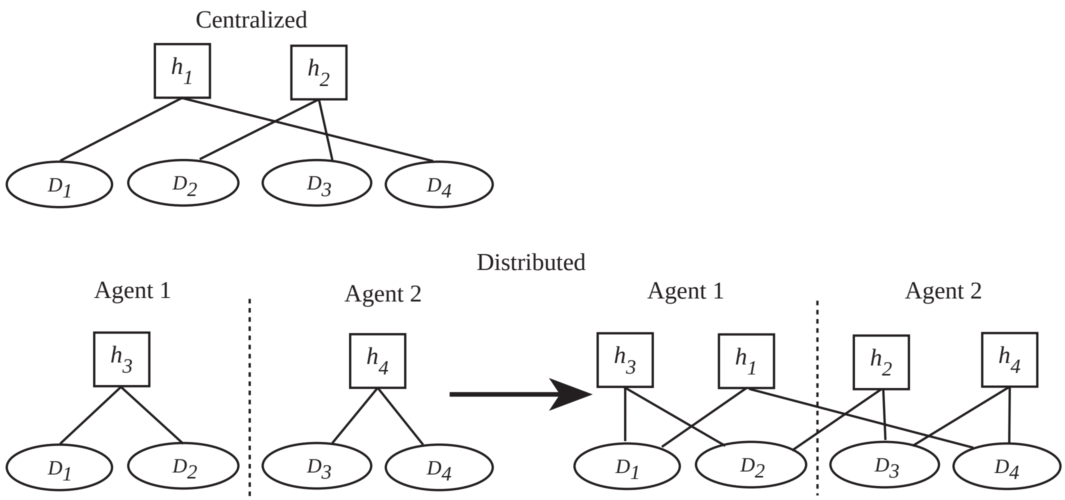

Figure 5: A counter example for the proof of theorem 2.
The centralized system, which has a view of the complete data set $D_1 \cup D_2 \cup D_3 \cup D_4$, initially chooses to construct hypotheses h_1 and h_2. If this is an acceptable explanation ($\widehat{BEL}(h_1)$ and $\widehat{BEL}(h_2) \geq AT$) no further propagation may be done. In the distributed case, the agents initially produce the hypotheses h_3 and h_4 (based on their own local data). In general, there will certainly exist data scenarios for which this interpretation would not be acceptable. However, even if the agents communicate and construct h_1 and h_2, $\widehat{BEL}(h_1)$ and $\widehat{BEL}(h_2)$ may be different than the approximate values computed in the centralized case, and could in fact lead to this alternative interpretation also being judged unacceptable.

Proof: Assume that the theorem is false—that $\forall \mathcal{D}$, $acceptable(\hat{\mathcal{I}}(C_{f_c}))$ implies $acceptable(\hat{\mathcal{I}}^G(C^G_{f_d}))$. It is easy to show the general existence of counter examples—see Figure 5. The reason that an acceptable centralized solution does not guarantee an acceptable distributed solution is that the division of data among the distributed agents may lead to the construction of a different set of hypotheses from that constructed by the centralized system. The construction of different hypotheses produces different belief approximations. This could cause the distributed system to produce a different acceptable solution or it could cause it to fail to find any acceptable solution. Even if the distributed system eventually constructs the same interpretation hypotheses as the centralized system, it may have done additional evidence propagation that causes the interpretation to be judged unacceptable. Furthermore, just because the distributed system may be driven to do more evidence propagation than the centralized system, this does not mean that its best solution is more likely to be the (true) global MPE (as long as its propagation is still incomplete).

Another way of stating theorem 2 is that when incomplete evidence propagation strategies are used, the probability of a distributed system producing an acceptable solution when an acceptable solution is produced by an equivalent centralized system may be less than one. This probability is a function of the propagation (control) strategy, the acceptance threshold, the characteristics of the domain, and the organization of agent interest areas. Understanding these relationships is a major focus of our current work. As a first step, we have explored how to decompose this probability so that we could determine bounds on its components based on reasonable probabilistic characterizations of a control strategy and domain.

So what we want to determine is: $P(acceptable(\hat{\mathcal{I}}^G(C^G_{f_d})) \mid acceptable(\hat{\mathcal{I}}(C_{f_c})))$, where C_{f_c} and $C^G_{f_d}$ are final contexts that meet the specified centralized and distributed termination criteria for the $\mathcal{D}$, respectively. (For simplicity, we will use $\hat{\mathcal{I}}^G$ for $\hat{\mathcal{I}}^G(C^G_{f_d})$ and $\hat{\mathcal{I}}^C$ for $\hat{\mathcal{I}}(C_{f_c})$ from this point on.) One way to decompose the probability is:

$$\begin{aligned}
&P(acceptable(\hat{\mathcal{I}}^G) \mid acceptable(\hat{\mathcal{I}}^C)) = \\
&\quad P(acceptable(\hat{\mathcal{I}}^G) \mid acceptable(\hat{\mathcal{I}}^C) \wedge \hat{\mathcal{I}}^G = \hat{\mathcal{I}}^C) * P(\hat{\mathcal{I}}^G = \hat{\mathcal{I}}^C \mid acceptable(\hat{\mathcal{I}}^C)) + \\
&\quad P(acceptable(\hat{\mathcal{I}}^G) \mid acceptable(\hat{\mathcal{I}}^C) \wedge \hat{\mathcal{I}}^G \neq \hat{\mathcal{I}}^C) * P(\hat{\mathcal{I}}^G \neq \hat{\mathcal{I}}^C \mid acceptable(\hat{\mathcal{I}}^C))
\end{aligned}$$

Here, we have decomposed the situation based on whether the distributed system finds that the most likely interpretation is the same interpretation as was found by the centralized system or not (remember that though the distributed system may have produced the same interpretation as the centralized system it may have done different propagation, so its belief ratings for the hypotheses of the interpretation may be different from those of the centralized system). Based on this decomposition, it is possible to identify some bounds on the component probabilities.

Consider the first conditional probability term in the decomposition:

$$\begin{aligned}
&P(acceptable(\hat{\mathcal{I}}^G) \mid acceptable(\hat{\mathcal{I}}^C) \wedge \hat{\mathcal{I}}^G = \hat{\mathcal{I}}^C) \geq \\
&\quad P(\forall h_i \in \hat{\mathcal{I}}^C, BEL(h_i) \geq AT \mid acceptable(\hat{\mathcal{I}}^C))
\end{aligned}$$

Proof: To prove this, we need to show that $acceptable(\hat{\mathcal{I}}^C) \wedge (\forall h_i \in \hat{\mathcal{I}}^C, BEL(h_i) \geq AT)$ implies $acceptable(\hat{\mathcal{I}}^G)$ (when $\hat{\mathcal{I}}^G = \hat{\mathcal{I}}^C$). Assume that this is false, so that $acceptable(\hat{\mathcal{I}}^C) \wedge (\forall h_i \in \hat{\mathcal{I}}^C, BEL(h_i) \geq AT) \wedge \neg acceptable(\hat{\mathcal{I}}^G)$. Based on the specified propagation strategy, $\neg acceptable(\hat{\mathcal{I}}^G)$ would cause the distributed system to do further evidence propagation until either $acceptable(\hat{\mathcal{I}}^G)$—which would contradict the assumption—or complete propagation has been done for all hypotheses in $\hat{\mathcal{I}}^G$ that are not acceptable—which also leads to a contradiction, as we will show. The complete propagation situation means that $\forall h_i \in \hat{\mathcal{I}}^G$ such that $\widehat{BEL}^G(h_i) < AT, \widehat{BEL}^G(h_i) = BEL^G(h_i)$. However, this contradicts the premise assumptions that $\forall h_i \in \hat{\mathcal{I}}^C, BEL(h_i) \geq AT$ since $\hat{\mathcal{I}}^G = \hat{\mathcal{I}}^C$ here.

$P(\forall h_i \in \hat{\mathcal{I}}^C, BEL(h_i) \geq AT \mid acceptable(\hat{\mathcal{I}}^C))$ is the probability that the hypotheses in the centralized interpretation are "acceptable" given these hypotheses true belief ratings whenever the partial propagation strategy finds that the hypotheses are acceptable given the approximate beliefs. This probability could be determined experimentally for domains in which complete propagation is at least possible (even if it is not possible for real-time problem solving), and could be experimentally estimated for domains in which complete propagation is absolutely intractable. It may also be possible to directly derive this probability from appropriate domain and strategy models. This would be easier than trying to derive $P(acceptable(\hat{\mathcal{I}}^G) \mid acceptable(\hat{\mathcal{I}}^C) \wedge \hat{\mathcal{I}}^G = \hat{\mathcal{I}}^C)$ since it would not require modeling the relationship between the organization of agent interest areas and the uncertainty/incompleteness of the locally available data, and how such conditions interact with a particular control strategy.

For the corresponding conditional probability term in which $\hat{\mathcal{I}}^G \neq \hat{\mathcal{I}}^C$, there is the following bound:

$$P(acceptable(\hat{\mathcal{I}}^G) \mid acceptable(\hat{\mathcal{I}}^C) \wedge \hat{\mathcal{I}}^G \neq \hat{\mathcal{I}}^C) \geq$$
$$P(\forall I_j (I_j \neq \hat{\mathcal{I}}^C) \Rightarrow (\forall h_i \in I_j, BEL(h_i) \geq AT) \mid acceptable(\hat{\mathcal{I}}^C) \wedge I_j \neq \hat{\mathcal{I}}^C)$$

Proof: The proof is analogous to the last proof.

$P(\forall I_j (I_j \neq \hat{\mathcal{I}}^C) \Rightarrow (\forall h_i \in I_j, BEL(h_i) \geq AT) \mid acceptable(\hat{\mathcal{I}}^C) \wedge I_j \neq \hat{\mathcal{I}}^C)$ is the probability that all interpretations of the data other than the most likely interpretation have hypotheses whose true belief ratings would make them "acceptable" interpretations, given that the most likely interpretation has been found to be acceptable using a partial propagation strategy. While this does provide a lower bound on the component conditional probability, it is a weak bound because it is unlikely to be very large for reasonable values of AT. We are working to improve this bound by exploring how different strategies affect the likelihood of locating acceptable interpretations when such interpretations exist. This will also assist in characterizing the relative likelihood that the distributed system will find the same or different interpretations than the centralized system—a key issue in assessing the target conditional probability.

6 Conclusions

In this paper we have taken a first step in analyzing the relationship between the solutions produced by a distributed interpretation system and those produced by an equivalent centralized system. We have shown that there are conditions under which it is possible to guarantee that a comparable solution is produced, and other conditions under which there is merely some probability of obtaining such a solution. We believe that this line of research will help our understanding of the assumptions that underlie the FA/C model of distributed problem solving and that it will assist in producing appropriate propagation and coordination strategies for distributed interpretation systems. It should also make it possible to analyze FA/C frameworks like DRESUN to determine whether the framework is effective in terms of making it possible to use all of the information that the (global) system has available or not (and what the effect there is in not being able to use all of this information).

We are currently working to extend our analyses in several directions: deriving specific probabilities of interest in connection with the theorems as a function of domain models and acceptance thresholds; understanding how to formalize the effects of different evidence propagation and coordination strategies; extending the analyses to deal with processing of only subsets of the available data; and generalizing to CDPS tasks other than sensor interpretation. For example, models of some interpretation domains show that a great deal of evidence is necessary to develop high belief ratings for hypotheses, so it should be unlikely that additional propagation would significantly change an acceptable interpretation's ratings if it already meets a high acceptance threshold. Thus, we expect that there should be some δ such that $acceptable(\hat{\mathcal{I}}C_{f_c})$ implies $acceptable'(\hat{\mathcal{I}}^G C^G_{f_d})$, where $acceptable'(\mathcal{I}C)$ means that $\forall h \in \mathcal{I}C, \widehat{BEL}(h) \geq AT - \delta$.

References

[1] Norman Carver and Victor Lesser, "A New Framework for Sensor Interpretation: Planning to Resolve Sources of Uncertainty," *Proceedings of AAAI-91*, 724–731, 1991.

[2] Norman Carver, Zarko Cvetanovic, and Victor Lesser, "Sophisticated Cooperation in FA/C Distributed Problem Solving Systems," *Proceedings of AAAI-91*, 191–198, 1991.

[3] Norman Carver, Victor Lesser, and Qiegang Long, "Resolving Global Inconsistency in Distributed Sensor Interpretation: Modeling Agent Interpretations in DRESUN," *Proceedings of the 12th International Workshop on Distributed Artificial Intelligence*, May, 1993.

[4] Gregory Courand, "Cooperation via Consensus Formation," *Proceedings of the 10th International Workshop on Distributed Artificial Intelligence*, 1990.

[5] Michael Huhns and David Bridgeland, "Multiagent Truth Maintenance," *IEEE Transactions on Systems, Man, and Cybernetics*, special issue on Distributed Artificial Intelligence, vol. 21, no. 6, 1437–1445, 1991.

[6] Victor Lesser and Daniel Corkill, "Functionally Accurate, Cooperative Distributed Systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 11, no. 1, 81–96, 1981.

[7] Victor Lesser and Daniel Corkill, "The Distributed Vehicle Monitoring Testbed: A Tool for Investigating Distributed Problem Solving Networks," *AI Magazine*, vol. 4, no. 3, 15–33, 1983 (also in *Blackboard Systems*, Robert Engelmore and Tony Morgan, editors, Addison-Wesley, 1988).

[8] Victor Lesser, "A Retrospective View of FA/C Distributed Problem Solving," *IEEE Transactions on Systems, Man, and Cybernetics*, special issue on Distributed Artificial Intelligence, vol. 21, no. 6, 1347–1362, 1991.

[9] Judea Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.

[10] Yun Peng and James Reggia, "Plausibility of Diagnositic Hypotheses: The Nature of Simplicity," Proceedings of AAAI-86, 140–145, 1986.