Bayesian Ontologies in Net-Centric Systems

Dr. Kathryn Blackmond Laskey
Center of Excellence for C4I
George Mason University
MS 4A6, 4400 University Drive
Fairfax, VA 22030-4400
klaskey@gmu.edu

Dr. Paulo Cesar G. da Costa
Center of Excellence for C4I
George Mason University
MS 4B5, 4400 University Drive
Fairfax, VA 22030-4400
pcosta@gmu.edu

Dr. Edward J. Wright
Information Extraction and Transport, Inc.
1911 N. Ft. Myer Drive, Suite 600
Arlington, VA 22209
ewright@iet.com

Abstract

In today's net-centric environment, there are new challenges to be met for any system to survive. Systems are required to fuse data from distributed, heterogeneous information sources operating asynchronously, to produce updated information, and to cope with diverse kinds of threats. Realizing this vision requires overcoming a number of technical challenges. Among these is the need for semantic interoperability among systems with different internal data models and vocabularies. Ontologies are seen as a key enabling technology for semantic interoperability, but current generation ontologies cannot represent or reason with incomplete data, a major shortcoming for Net-Centric systems. This paper proposes the use of probabilistic ontologies within a service-oriented architecture as a means to enable semantic interoperability in net-centric systems

Keywords: Bayesian ontologies, probabilistic reasoning, MEBN, PR-OWL, uncertainty, geospatial systems, decision support, SOA.

1. Introduction

The dawn of the new millennium has brought profound changes in the way wars are fought. Established doctrines and tactics are increasingly giving way to new trends in warfare. Big wars against monolithic enemies employing rigid doctrines have given way to asymmetric conflict against unpredictable opponents. We now face terrorist groups instead of armies, guerrilla tactics instead of open combat maneuvers, networked cells embedded in civilian populations instead of troops entrenched in the battlefield, and loosely connected units instead of a structured hierarchy of command.

Confronting change of this magnitude is an enormous challenge to a military force that has grown accustomed to dominance by overwhelming force. We are coming to the realization that even the strongest army is vulnerable unless its tactics and strategies are designed to cope with the enemy's modus operandi. Traditional combat gave rise to stovepiped systems – crafted for use by single organizations, for single purposes, and relying upon idiosyncratic database schema and input-output formats. Because stovepiped systems were not designed for interoperability, they require labor-intensive manual transformation of outputs for use by other systems. Such systems are very effective against the enemy they were designed to counter, and have provided the West a clear dominance in information technology.

However, the world has changed. The time is ripe to profit on the situation by taking the opportunity to improve. To use an overstressed analogy, just as Goliath's brute force did not prevail against David's tactical surprise, systems designed for another era of warfare are vulnerable against smaller, more agile enemies. Superior power is a major asset, but it is not enough. More must be done to better understand the enemy and to react proactively. In the wake of the 2001 World Trade Center bombings, security experts have noted[1] that the linchpin of the homeland security system is not Jersey barriers and metal detectors, but intelligence that a threat may be coming. Getting the edge over asymmetric threats requires lighter, faster, more flexible systems that can interoperate seamlessly.

In response to this need there is the idea of a Net-Centric World, in which autonomous software agents interoperate seamlessly, sharing information and behaving collaboratively. In principle, Net-Centric systems can be vastly superior to current complex, do-it-all systems. Distributed solutions in which each agent has timely access to mission-critical information can provide timely information where and when it is needed. Systems to survive in this environment need to be designed for easy updating and synchronization, for handling data from disparate sources and fusing it into mission-relevant knowledge, and for coping with multi-level security protocols to enable information to be accessed as needed while preventing non-authorized use.

The above is an impressive list of technical requirements, the complexity of which is often underestimated. A recurring misconception is the "bandwidth fallacy": that providing the bandwidth to move massive volumes of data automatically provides information security. Although a tempting idea, it doesn't survive to the first reality check. With present day technology, the transformation from data to knowledge still requires extensive labor-intensive human intervention. Thus, no matter how much bandwidth a system gets, there will always be the knowledge bottleneck.

This paper addresses a combination of techniques devised to bring the vision of a Net-Centric World closer to realization. In section two, we present Web Services as an enabler for achieving interoperable systems, the need of including semantics to Web Services protocols, and a case study to illustrate how semantics are beneficial to system interoperability. Then, in section three we present ontologies as a more sophisticated means of conveying semantics, argue that semantically aware systems must be able to represent and process uncertainty, and present probabilistic ontologies as a solution. Section four is devoted to the backbone technologies that allow such concepts to be realized, Multi-Entity Bayesian Networks and Probabilistic OWL. The final section is a brief discussion of the issues raised earlier in the paper.

2. Web Services: Enabling Interoperability

2.1. The P-F-B Triangle

Figure 1 depicts a typical Web Services scenario, which makes use of the publish-find-bind triangle to match service providers with consumers. A service provider publishes a service description. A consumer searches a service registry for a service satisfying his criteria, analyzes the included information (or link to information) about the message structure to be exchanged and the address to exchange it, and interacts with the service to retrieve the resources needed. The discovery mechanism identifies a provider

[1] Washington Post, September 12, 2001.

matching the consumer's criteria, and the binding mechanism makes the connection that enables the exchange to take place.

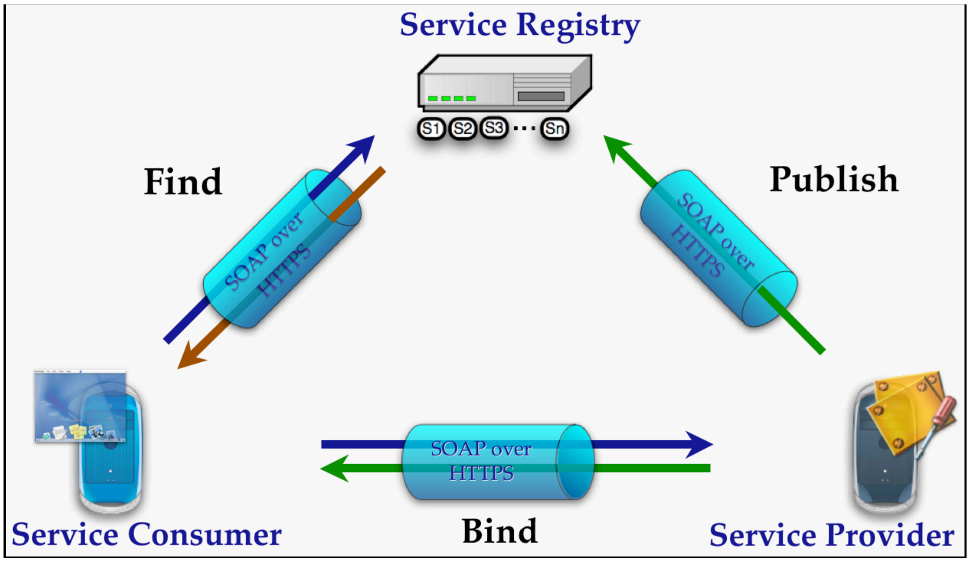

Figure 1 – The P-F-B Triangle

In this triangle, there are implicit, unspoken challenges for which a principled representation of uncertainty is needed. For example:

- The publisher has to choose vocabulary with which to describe the service (or some other resource related to the service). The vocabulary sets the properties for that class of item. Service ontology developers attempt to define the "right" set and structure of properties for the anticipated users. The consumer must know and understand the semantics of the chosen property vocabulary because these are the properties used to describe the class and its instances, and the consumer must understand and use the same vocabulary or there must be a known and accessible mapping between the properties used for description and those used as search categories. There are many opportunities for uncertainty about intended meanings.
- The publisher uses the chosen property vocabulary as the basis to describe and register instances of that class. This means that the publisher associates values with the properties and registers the instance. But what is the vocabulary for the values? All parties may agree that something has the property *color* and on the meaning of that property, but if the publisher uses only primary colors and the subscriber's search criterion asks for the color *pink,* the latter will never find a match for items the first had catalogued. How does a client's requested value relate to a provider's published values? Do they agree on the vocabulary? Do they agree on the mechanism to mediate vocabulary mismatches?
- The publisher chooses a property vocabulary and creates instance descriptions by associating values. One can infer what properties the publisher considers important by which properties s/he chooses to populate, assuming values are not necessarily assigned for all possible properties. But what of the consumer's priorities when assigning search criteria? If the consumer assigns relative importance, how does the search engine trade off among different combinations of matches across the consumer's search criteria, and how are missing attribute values handled?

Beyond publish-find-bind for a single, simple service, the Net-Centric vision is requires services to be provided at the appropriate level of granularity, combining atomic services into more complex tasks.

Web Services is the most prevalent implementation of the more general Service Oriented Architecture (SOA). SOA has become the leading approach for accessing and using distributed resources developed by independent entities and working with independently developed vocabularies and associated semantics. The advent of SOA marks a transformation from applications running in an isolated environment, with little interaction between requesters and providers of information, into one in which information and other resources are accessed and used in a much more dynamic, interactive, and unpredictable fashion.

2.2. Semantics in Web Services

The supporting technology for the SOA model is composed of XML-based standards and protocols focused on providing a shared understanding of the available services. Currently, accepted standards for developing solutions based on Web Services include SOAP, a message structure used for exchanging XML serializations of content and message handling instructions in a decentralized, distributed environment [1], and the Web Services Description Language (WSDL), which represents messages exchanged when invoking a Web Service [2].

However, these XML-based structures do not have the ability to explicitly formalize the underlying semantics of a given Web Service description, rendering them insufficient to ensure a common understanding of the described Web Service. As pointed out by Paolucci et al. [3], two identical XML descriptions may have different meanings depending on who uses them and when. Because it is unrealistic to expect that all providers and consumers will have equivalent perspectives and knowledge regarding a given service, a common understanding of a given Web Service can be reached only at the semantic level, where the different perspectives and knowledge can be matched.

Not surprisingly, the need for semantic-aware resource descriptions is widely recognized, and is being addressed by research focused on enabling Web Service providers to describe the properties and capabilities of their Web Services in unambiguous, computer-interpretable form (e.g., OWL-S [4], WSMO [5], SWSL [6], SAWSDL [7], and WSDL-S [8]).

The syntactical interoperability achieved by current SOA standards implies that applications can process each other's data formats. As an example, two systems could agree that a given field stores a floating-point number in a standard floating-point format, and then exchange the number 3.2 stored in that field. All that can be extracted from this exchange is the datatype of 3.2, with no regard to what the information means.

On the other hand, semantic interoperability goes further in requiring applications to have the same interpretation of that data. In the previous example, the floating point number 3.2 might represent the processor speed of a computer. The ability to represent and exchange this semantic information is a much stronger requirement than simple type consistency.

In stovepiped systems, semantics are in the mind of humans. Humans must resort to catalogs and other natural language documentation to interpret the meaning of system outputs, and must perform labor-intensive, manual transformation to interchange data between systems. In Net-Centric systems, semantics are embedded in the code, and require formal, machine-interpretable representations. While stovepipe systems would require only syntactical interoperability, Net-Centric systems would need semantic information to be embedded in service descriptions, as a means for enabling consumers and providers to have a common understanding of issues such as:

- What does the service do?
- What inputs does it require and what results does it produce?
- What are conditions (constraints/policies) for use?
- How to invoke it? (Address & WSDL description)

However, adding semantics to Web Services descriptions solves only part of the problem, as it provides a means for consumers and providers to exchange more expressive information, instead of merely agreeing on data types. Still, as we discuss in section three, more is needed to foster that common understanding, and a comprehensive description of the knowledge to be exchanged is necessary.

2.3. A Case Study: Geospatial Services

The focal point of the battlefield command post is the map. Through interactions with the map, the commander and staff collaborate to build a common operating picture. This common operating picture displays the area of operations, the militarily significant features of the terrain, the locations of adversary and friendly forces, and the evolving plan. A generation ago, planning centered on a paper map, its overlays of acetate covered with marks of grease pencils wielded by the staff members congregated around it. The manual geospatial analysis process was tedious, time-intensive, error-prone, and difficult to share.

Today, the paper map has been replaced in brigade and higher headquarters with a digitized map projected onto a large-screen display. The grease pencil has become an input device for drawing objects or selecting pre-computed overlays from a menu of options. The map and overlays are stored in the computer as data structures, are processed by algorithms that can generate in seconds products it would take soldiers many hours of tedious effort to duplicate, and can be sent instantly to relevant consumers anywhere on the Global Information Grid (GIG), the information processing infrastructure of the United States Department of Defense (DoD). The GIG is the physical infrastructure to enable Network-Centric Operations, the DoD's new doctrine for warfare in the 21st Century.

Advanced automated geospatial tools (AAGTs) transform commercial geographic information systems (GIS) into useful military services for Network Centric Operations. Because of their basis in commercial GIS, they also have widespread applicability to fire, police, disaster relief, and other problems characterized by a command hierarchy. The advanced situation awareness provided by AAGTs can do much more than simply speed up calculations. They are changing the way military operations are conducted, reducing time to produce analysis, avoiding rework, and reducing bandwidth by sending results instead of raw data.

The development of tools is shaped by military necessity, but as the new century dawns, the decision making process itself is being shaped by the automated tools that provide warfighters with more robust situational awareness. One example of a tool designed to support the development of geospatial systems is the Geospatial Battle Management Language (geoBML) [9] for terrain reasoning, which provides a common representation of military mission concepts that are suitable for automated processing.

GeoBML has three distinct tiers of spatial objects (SO). Tier 1 constitutes those SO's which are based exclusively or primarily upon the terrain and can be pre-computed without being informed by the other factors of METT-TC. Examples include:

- Cross County Mobility; Obstacle; Cover and Concealment SO's

- Maneuver Networks
- Mobility and Choke Point SO's
- Fields of Fire and Dead Space SO's

Tier 2 SO's are those that might support certain missions or tactical tasks, but are based upon a strong set of terrain attributes and for which candidate SO's can be pre-computed. Examples are:

- Indirect Fire Firing Positions
- Assembly Areas
- Battle Positions
- Engagement Areas

Finally, Tier 3 SO's are very specific objects that have been selected to support a specific Course of Action (COA) and are associated with a plan or order. In many cases they have been chosen from the Tier 2 candidate SO's and been further refined based upon METT-TC. They may also include graphic control measures and other items that are often associated with or influence the perception of terrain.

3. Representing Semantics and Uncertainty

3.1. Ontologies

The term Ontology was borrowed from philosophy. Its roots can be traced back to Aristotle's metaphysical studies of the nature of being and knowing[2]. Nonetheless, use of the term ontology in the information systems domain is relatively new, with the first appearance occurring in 1967 [10, page 22].

One can find many different definitions for the concept of ontology applied to information systems, each emphasizing a specific aspect its author judged as being more important. For instance, Gruber [11] defines an ontology as a formal specification of a conceptualization or, in other words, a declarative representation of knowledge relevant to a particular domain. Uschold and Gruninger [12] define an ontology as a shared understanding of some domain of interest. Sowa [13, page 492] defines an ontology as a product of a study of things that exist or may exist in some domain.

With so many possibilities for defining what an ontology is, one way of avoiding ambiguity is to focus on the objectives being sought when using it. For the purposes of the present research effort, the most important aspect of ontologies is their role as a structured form of knowledge representation. Thus, our definition of ontologies is a pragmatic one that emphasizes the purposes for which ontologies are used:

Definition 1 [from 14]: An ontology is an explicit, formal representation of knowledge about a domain of application. This includes:

a) Types of entities that exist in the domain;

b) Properties of those entities;

c) Relationships among entities;

d) Processes and events that happen with those entities;

[2] The term metaphysics means beyond the study of physics

where the term entity refers to any concept (real or fictitious, concrete or abstract) that can be described and reasoned about within the domain of application. Ontologies are used for the purpose of comprehensively describing knowledge about a domain in a structured and sharable way, ideally in a format that can be read and processed by a computer.

In a hypothetical Geospatial Ontology, examples of types would include mobility corridors, vehicles, etc., examples of attributes would include the width of a mobility corridor, the maximum speed of a vehicle, etc., and examples of relationships include the fact that vehicles travel on mobility corridors. It is important to note that such concepts must be represented in a way that is formal enough to allow reasoning.

3.2. More on Geospatial Data: Example of Uncertainty Representation

The example below, extracted from [15], illustrates the challenges and opportunities of uncertainty management in geospatial information systems via a Cross Country Mobility (CCM) analysis. CCM analysis is performed to evaluate the feasibility and desirability of enemy and friendly courses of action. The Cross Country Mobility (CCM) Tactical Decision Aid (TDA) predicts the speed that a specific military vehicle or unit can move across country (off roads) based on the terrain. The terrain factors that influence CCM speed are slope, soil type, soil wetness, vegetation and vegetation attributes, ground or surface roughness, and presence of obstacles.

There are several CCM analysis models commonly in use by military organizations in the U.S. and around the world. The CCM product of Figure 2 was produced using the DMA CCM algorithm (DMS, 1993). CCM products can be generated for specific vehicle types, for classes of vehicles, or for military unit types. The products can be used as inputs to algorithms for producing mobility corridors, or combined with other information to generate avenues of approach for friendly or enemy forces. Traditional CCM algorithms use point estimates of their input data and produce point estimates of predicted speeds. Traditional CCM displays show predicted speeds without any attempt to estimate or communicate the quality of the prediction based on the quality of the underlying data and the quality of the algorithm used to make the prediction.

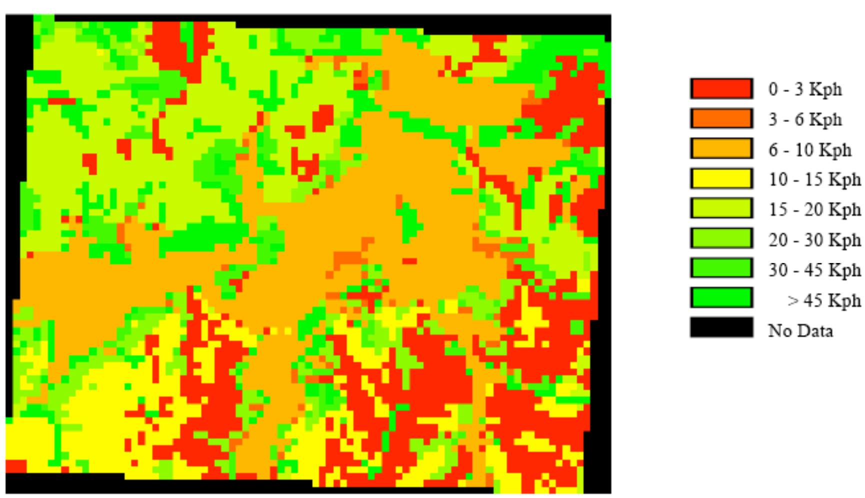

Figure 2 - Traditional CCM Product
(M1 Tank, DMA Mobility Model, ITD Data, Korea)

There are many sources of uncertainty in CCM estimates. Input data on the factors that influence speed may be contain errors. In many cases, the input parameters required by models may be unavailable, and must be estimated using a combination of auxiliary models and human judgment. Models for predicting speed from input parameters are imperfect. Uncertainty can have decision implications, and decision making can be improved by properly considering uncertainty in decision support algorithms.

A major component for decision-making on geospatial data is the visualization of uncertainty, which is essential to communicate uncertainties to decision makers. This helps to prevent decision makers from being blinded by the quality of the display, and to make them aware of the underlying uncertainty of the product. Figure 3 shows a visualization of CCM with associated uncertainty.

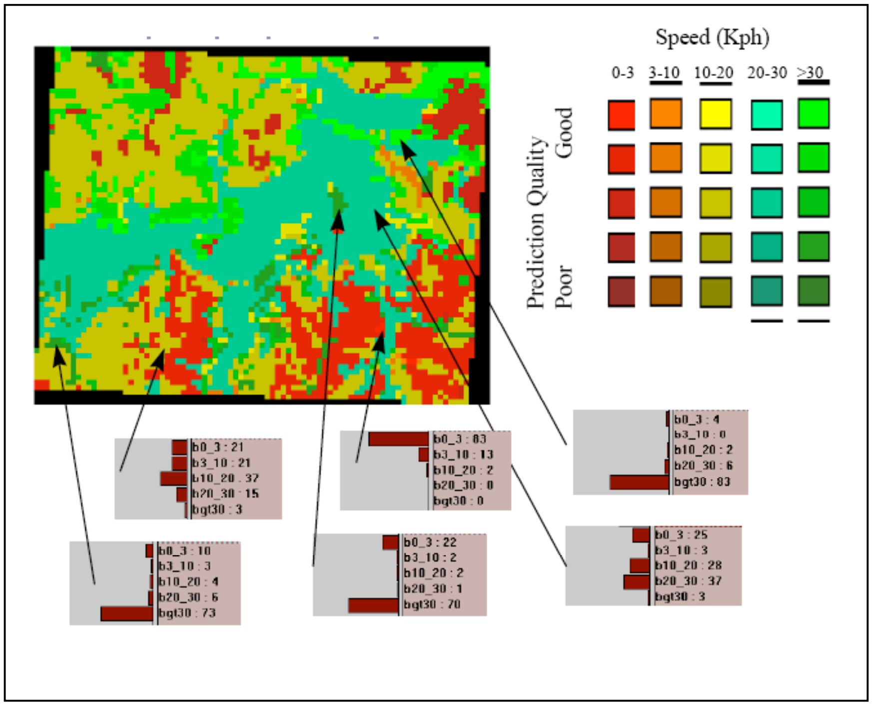

Figure 3 - CCM Product with Visualization of Uncertainty in the CCM Prediction

CCM uncertainty is shown two ways, in the legend and via interactive histograms that the user can control. The legend is a bi-variate legend where the color represents the predicted CCM speed range, and the quality of the color represents the quality of the prediction. There is enough information in the legend that it is difficult to interpret the product colors. This difficulty is exacerbated by the difficulty of matching colors from computer monitor to printed hardcopy. To offset the difficulty in interpretation, user controlled popup histograms were provided on the digital display.

This CCM display provides more information to decision makers about the quality of the prediction and (in the interactive versions) the popup histograms provide a means to query for more detailed predictions at specific points. The model also makes it possible to run simulation experiments assessing the results for (say) 100 randomly assigned routes, thus providing added situational awareness, resulting in better decisions and fewer mission failures. This brief example demonstrates the importance of representing, properly managing, and communicating to decision makers information about uncertainty in the GIS products used for military planning

3.3. Probabilistic Ontologies

Ontologies comprise a common set of terms for describing and representing a domain in a way that allows automated tools to use the stored data in a more context-aware fashion, intelligent software agents to perform better knowledge management, and many other benefits achieved by a standardized, intensive use of metadata. However, they failed to provide a standardized means to convey both the structural and numerical information required to represent and reason with uncertainty in a principled way. Probabilistic Ontologies (PO), on the other hand, are designed for comprehensively describing knowledge about a domain and the uncertainty associated with that knowledge in a principled, structured and sharable way. Therefore, POs provide a coherent representation of statistical regularities and uncertain evidence, an ideal way of representing and propagating uncertainty in geospatial systems.

Following the ongoing example, consider the need of aggregating geospatial information from several databases. Suppose we consult three different databases, all three of which label a particular area as forested. Each report is tagged with a particular credibility. Because the three reports agree, standard statistical aggregation technologies would label the region as forested and assign a higher credibility than the three individual credibilities. However, suppose that all three databases obtained their raw data for this area from the same satellite image, and all three applied similar algorithms for assigning a ground cover type label. In this situation, the credibility of the aggregate report is no greater than any of the individual input credibility values. In this case, we need to represent not just a single credibility number, but dependency information about how the credibility depends on the sensor and the data processing algorithm.

The example above highlights a major distinction between ontologies and POs. If using the former approach, in order to satisfy the representational nuances depicted here one would have to annotate a standard ontology with numerical probabilities, while also being forced to devise ad-hoc means to convey the structural constrains and dependencies not represented by the numbers. POs, on the other hand, provide a flexible way for expressing such domain subtleties, a crucial requirement for dealing with uncertainty in geospatial systems. Indeed, if the systems providing input give no data quality information, or supply insufficient information for the probabilistic reasoner to determine unambiguously the structure and/or probabilities for the Bayesian network, then the fusion system has an additional inference challenge – to determine the appropriate BN for fusing the diverse inputs.

Intuitively, an ontology that has probabilities attached to some of its elements would qualify for this label, but such a limited definition is inadequate for our purposes. Merely adding probabilities to concepts does not guarantee interoperability with other ontologies that also carry probabilities. More is needed than syntax for including probabilities if we are to justify a new category of ontologies. Definition 2 below expands the definition of ontology presented before to allow for a principled representation and reasoning with incomplete data:

Definition 2 [from 14]: A probabilistic ontology is an explicit, formal representation of knowledge about a domain of application. This includes:

a) Types of entities that exist in the domain;

b) Properties of those entities;

c) Relationships among entities;

d) Processes and events that happen with those entities;

e) Statistical regularities that characterize the domain;

f) Inconclusive, ambiguous, incomplete, unreliable, and dissonant knowledge related to entities of the domaind,; and,

g) Uncertainty about all the above forms of knowledge,

where the term entity refers to any concept (real or fictitious, concrete or abstract) that can be described and reasoned about within the domain of application. Probabilistic Ontologies are used for the purpose of comprehensively describing knowledge about a domain and the uncertainty regarding that knowledge in a principled, structured and sharable way, ideally in a format that can be read and processed by a computer.

4. Technology Enablers to Interoperability

4.1. Multi-Entity Bayesian Networks (MEBN)

MEBN logic [16] combines Bayesian probability theory with classical First Order Logic and allows for the representation of Bayesian models with repeated structure. Probabilistic knowledge is expressed as a set of MEBN fragments (MFrags) organized into MEBN Theories. An MFrag is a knowledge structure that represents probabilistic knowledge about a collection of related hypotheses. Hypotheses in an MFrag may be *context* (must be satisfied for the probability definitions to apply), *input* (probabilities are defined in other MFrags), or *resident* (probabilities defined in the MFrag itself). An MFrag can be instantiated to create as many instances of the hypotheses as needed (e.g., an instance of the “Disease” hypothesis for each patient at a clinic). Instances of different MFrags may be combined to form complex probability models for specific situations. A MEBN theory is a collection of MFrags that satisfies consistency constraints ensuring the existence of a unique joint probability distribution over instances of the hypotheses in its MFrags.

MEBN inference begins when a query is posed to assess the degree of belief in a target random variable given a set of evidence random variables. We start with a generative MTheory, add a set of finding MFrags representing problem-specific information, and specify the target nodes for our query. The first step in MEBN inference is to construct a situation-specific Bayesian network (SSBN), which is a Bayesian network constructed by creating and combining instances of the MFrags in the generative MTheory. When each MFrag is instantiated, instances of its random variables are created to represent known background information, observed evidence, and queries of interest to the decision maker. If there are any random variables with undefined distributions, then the algorithm proceeds by instantiating their respective home MFrags. The process of retrieving and instantiating MFrags continues until there are no remaining random variables having either undefined distributions or unknown values. A SSBN may contain any number of instances of each MFrag, depending on the number of entities and their interrelationships. Next, a standard Bayesian network inference algorithm is applied. Finally, the answer to the query is obtained by inspecting the posterior probabilities of the target nodes.

4.2. Probabilistic Web Ontology Language (PR-OWL)

PR-OWL is a MEBN-based extension to the OWL ontology language. It represents MEBN Theories in a XML-compliant format. It is open-source, freely available solution for representing knowledge and associated uncertainty in a principled way. In this initial phase, PR-OWL is an upper ontology to guide the development of probabilistic ontologies. Daconta *et al.* define an upper ontology as a set of integrated ontologies that characterizes a set of basic commonsense knowledge notions [17]. Currently, these basic commonsense notions are related to representing uncertainty in a principled way using OWL syntax. If PR-OWL were to become a W3C Recommendation, this collection of notions would be formally incorporated into the OWL language as a set of constructs that can be employed to build probabilistic ontologies.

The PR-OWL upper ontology for probabilistic systems consists of a set of classes, subclasses and properties that collectively form a framework for building probabilistic ontologies. The first step toward building a probabilistic ontology in compliance with our definition is to import into any OWL editor an OWL file containing the PR-OWL classes, subclasses, and properties.

From our definition, it is clear that nothing prevents a probabilistic ontology from being "partially probabilistic". That is, a knowledge engineer can choose the concepts he/she wants to include in the "probabilistic part" of the ontology, while writing the other concepts in standard OWL. In this case, the "probabilistic part" refers to the concepts written using PR-OWL definitions and that collectively form a MEBN Theory. There is no need for all the concepts in a probabilistic ontology to be probabilistic, but at least some have to form a valid MEBN Theory. Of course, only the concepts that are part of the MEBN Theory will be subject to the advantages of the probabilistic ontology over a deterministic one.

The subtlety here is that legacy OWL ontologies can be upgraded to probabilistic ontologies only with respect to concepts for which the modeler wants to have uncertainty represented in a principled manner, make plausible inferences from that uncertain evidence, or to learn its parameters from incoming data via Bayesian learning. While the first two are direct consequences of using a probabilistic knowledge representation, the latter is a specific advantage of the Bayesian paradigm, where learning falls into the same conceptual framework as knowledge representation.

4.3. Discovery with Probabilistic Ontology

Figure 4 shows a simplified scheme for SOA using probabilistic semantic mapping. As a means to illustrate this scheme, we will devise fictitious examples involving Web Service providers within the geospatial reasoning domain. In this scheme, a service consumer or provider that conveys semantic information (ontology that it abides to, metadata about its requests, parameters, etc.) is called a SOA node Level 1, whereas a SOA node that has no semantic awareness is called a SOA node Level 0.

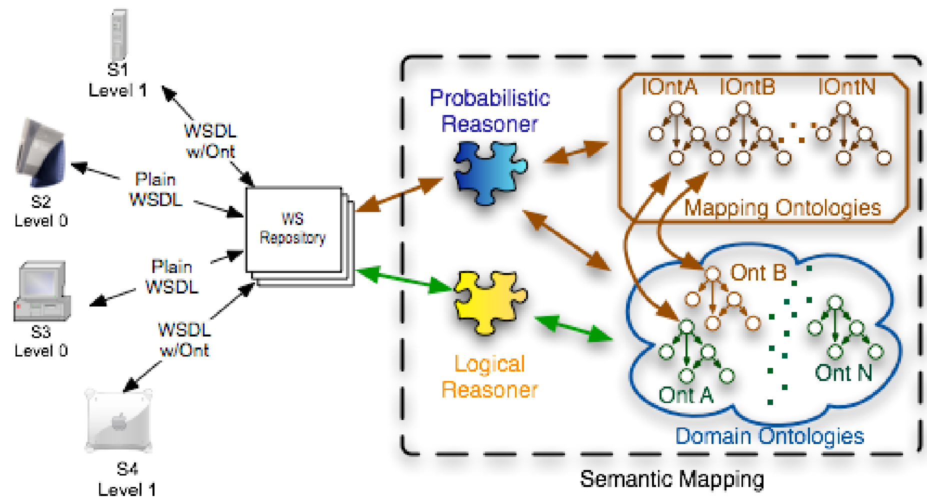

Figure 4 - Probabilistic Semantic Mapping for Web Services

In our first use case, S1 needs to generate a travel plan and requests a service for assessing the possibility of flooding in a given region due to recent heavy rains. Being a Level 1 client, S1 sends its request with embedded data about the ontology it references and other semantic information regarding its request (e.g. coordinate system used, expected QoS, etc.). The WS repository, which itself uses an ontology, finds S4, another Level 1 client using the same ontology as S1. This ontology is the PR-OWL ontology "OntB", which represents a probabilistic model of the geospatial domain and has the ability to perform a probabilistic assessment of the requested information. In this case, the request was probabilistic, but the uncertainty involved was related to the service itself (a probabilistic query on a uncertainty-laden domain), and not to the service exchanging process. In other words, the exchange was completed using the logical reasoner alone, since there was a perfect matching in terms of ontologies (both S1 and S4 abide to the same PR-OWL ontology) and the parameters of the requested service, and thus no probabilistic mapping was necessary. (yet, note that S1's query made use of OntB's ability to represent uncertainty about the geospatial domain.)

In a variation of the previous case, let's suppose that no perfect match between the request and the available providers is found. In this case, the probabilistic reasoner accesses the WS repository to search for the most suitable service given the parameters of S1's request. During that process, it analyses the mapping ontologies related to "OntB" (the ontology referenced by S1) and the domain ontologies related to the services it deemed promising to fit S1's request. In the end, an ordered list of possible providers is built, and the best possible answers will be returned to S1. This simple example shows that there might be many combinations of the use of logical and probabilistic reasoners and ontologies to match the needs of a specific request.

5. Discussion

The Net-Centric vision require explicit semantics as an enabling factor for interoperable systems. The ability to represent and reason with uncertain, incomplete data is a major requirement to military systems, and the lack of such capacity affects mission performance by ignoring an important part of situational awareness. Annotating a standard ontology with "uncertainty attributes" is not enough to fill this gap, and a rich representation of uncertainty is needed.

Probabilistic ontologies provide a principled representation of uncertainty in a given domain, and we have shown some of its uses in extending the reach of Web Services as a technology for exchanging knowledge. Although the concept of a semantic-enabled SOA is in its infancy, we believe much can be achieved by employing both complete and incomplete knowledge to optimize the way resources are exchanged.

References

[1] Mitra, N. "SOAP Version 1.2 Part 0: Primer", W3C Recommendation 24 June 2003, available at http://www.w3.org/TR/soap12-part0/

[2] Chinnici, R.; Moreau, J.; Ryman, A.; and Weerawarana, S. "Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language", W3C Candidate Recommendation 27 March 2006. Available at http://www.w3.org/TR/wsdl20

[3] Paolucci, M.; Kawamura, T.; Payne, T.; and Sycara, K. (2002) "Importing the Semantic Web in UDDI". In C. Bussler, R. Hull, S. McIlraith, M. Orlowska, B. Pernici, and J. Yang, editors, Proceedings of the Workshop on Web Services, E-Business and Semantic Web, CAiSE. Toronto, Canada.

[4] Martin, D. (ed.) "OWL-S: Semantic Markup for Web Services". Available at http://www.daml.org/services/owl-s/1.1/overview/

[5] Domingue, J.; Lausen, H., and Fensel, D. (Chairs) ESSI Web Services Modeling Ontology Working Group. Website, available at http://www.wsmo.org.

[6] Battle, S. *et al.* (2005) "Semantic Web Services Framework (SWSF) Overview". Available at http://www.daml.org/services/swsf/1.0/overview/

[7] W3C (2006) Semantic Annotations for WSDL Working Group. Website, available at http://www.w3.org/2002/ws/sawsdl.

[8] Web Service Semantics - WSDL-S, W3C Member Submission 7 November 2005, available at http://www.w3.org/Submission/WSDL-S/#S42.

[9] Hieb, M.; Powers, M.; Pullen, M.; and Kleiner, M. (2006) "A Geospatial Battle Management Language (geoBML) for Terrain Reasoning." In proceedings of the 11th International Command and Control Research and Technology Sysposium. Cambridge: UK.

[10] Smith, B. (2004). Ontology: Philosophical and Computational. Retrieved January 15, 2005, from http://ontology.buffalo.edu/smith//articles/ontologies.htm.

[11] Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2), 199-220.

[12] Uschold, M.; Gruninger, M. (1996). Ontologies: Principles, Methods, and Applications. Knowledge Engineering Review, 11(2), 93-155.

[13] Sowa, J. (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundations. Pacific Grove, CA, USA: Brooks/Cole.

[14] Costa, P. (2005). Bayesian Semantics for the Semantic Web. Doctoral Thesis, School of Information Technology and Engineering, George Mason University. Fairfax, VA, USA. Available at http://hdl.handle.net/1920/455.

[15] Wright, E. (2002). Understanding and Managing Uncertainty in Geospatial Data for Tactical Decision Aids. Doctoral Dissertation, School of Computational Sciences, George Mason University: Fairfax, VA, USA.

[16] DMS, (1993) "Procedural Guide for Preparation of DMA Cross-Country Movement (CCM) Overlays", Defense Mapping School Student Text DMS ST 330, Ft. Belvoir, VA.

[17] Laskey, K. MEBN: A Logic for Open-World Probabilistic Reasoning. Department of Systems Engineering and Operations Research, George Mason University, Fairfax, VA, USA. Avail-able at http://hdl.handle.net/1920/461.

[18] Daconta, M., Obrst, L.; and Smith, K. (2003) The Sematic Web: A guide to the future of XML, Web services, and knowledge management. Indianapolis, IN: Wiley Publish-ing, Inc. 312.