PR-OWL 2 Case Study: A Maritime Domain Probabilistic Ontology

Kathryn Blackmond Laskey, Richard Haberlin, Paulo Costa
Volgenau School of Engineering
George Mason University
Fairfax, VA USA
[klaskey, rhaberli, pcosta]@gmu.edu

Rommel Novaes Carvalho
Brazilian Office of the Comptroller General
Brasília, Brazil
rommel.carvalho@gmail.com

***Abstract*—Probabilistic ontologies incorporate uncertain and incomplete information into domain ontologies, allowing uncertainty in attributes of and relationships among domain entities to be represented in a consistent and coherent manner. The probabilistic ontology language PR-OWL provides OWL constructs for representing multi-entity Bayesian network (MEBN) theories. Although compatibility with OWL was a major design goal of PR-OWL, the initial version fell short in several important respects. These shortcomings are addressed by the latest version, PR-OWL 2. This paper provides an overview of the new features of PR-OWL 2 and presents a case study of a probabilistic ontology in the maritime domain. The case study describes the process of constructing a PR-OWL 2 ontology using an existing OWL ontology as a starting point.**

Keywords- Probabilistic ontology, Multi-Entity Bayesian networks, PR-OWL, OWL, Maritime domain ontology, Uncertainty Modeling Process for Semantic Technologies

I. INTRODUCTION

The emphasis on net-centric operations and the shift to asymmetric warfare have created new challenges for automated information integration. To meet these challenges, developers are recognizing the need to combine explicit representation of domain semantics with the ability to represent and reason with uncertainty. Probabilistic ontologies allow the representation of uncertainty about attributes of and relationships among domain entities. Probabilistic OWL (PR-OWL) [1] is an OWL upper ontology for representing probabilistic ontologies. Compatibility with OWL was a major design goal for PR-OWL. However, the initial release of PR-OWL falls short of complete compatibility in several important respects. First, there is no mapping in PR-OWL to properties of OWL. Second, although PR-OWL has the concept of meta-entities, which allows the definition of complex types, it lacks compatibility with existing types already present in OWL. These problems have been noted in the literature [2]:

> PR-OWL does not provide a proper integration of the formalism of MEBN and the logical basis of OWL on the meta level. More specifically, as the connection between a statement in PR-OWL and a statement in OWL is not formalized, it is unclear how to perform the integration of ontologies that contain statements of both formalisms.

Carvalho [3] proposed a new syntax and semantics, defined as PR-OWL 2, which improves compatibility between PR-OWL and OWL in two important respects. First, PR-OWL 2 follows the approach suggested by Poole et al. to formalizing the association between random variables from probabilistic theories with the individuals, classes and properties from ontological languages such as OWL. Second, PR-OWL 2 allows values of random variables to range over OWL datatypes.

This paper presents an overview of PR-OWL 2, describes the key features that improve compatibility with OWL, discusses an open-source tool for building PR-OWL 2 probabilistic ontologies, and describes a use case of a PR-OWL 2 ontology for maritime domain awareness.

II. A PROBABILISTIC ONTOLOGY IN PR-OWL

A. *PR-OWL 1: An Upper Ontology for MEBN Theories*

PR-OWL provides constructs to define probabilistic ontologies in the OWL ontology language. The initial version, PR-OWL 1, is an OWL upper ontology for representing MEBN theories [4]. MEBN is a first-order probabilistic language (FOPL) [5] that allows probabilities to be assigned in a consistent way to logical statements. MEBN represents the world as entities that have attributes and are related to other entities. Knowledge about the attributes of entities and their relationships to each other is represented as a collection of MEBN fragments (MFrags) organized into MEBN Theories (MTheories). An MFrag represents a conditional probability distribution for instances of its resident random variables given their parents in the fragment graph and the context nodes. An MTheory is a set of MFrags that collectively satisfies consistency constraints ensuring the existence of a unique joint probability distribution over instances of the random variables represented in each of the MFrags within the set. A PR-OWL ontology encodes domain knowledge as a set of MFrags. A PR-OWL reasoner uses the probability information encoded in the MFrags to compute responses to probabilistic queries.

This research was partially supported by the Office of Naval Research (ONR), under Contract N00173-09-C-4008. Rommel Carvalho would like to thank the Brazilian Office of the Comptroller General (CGU) for their active support since 2008 and for providing the human resources necessary to conduct this research.

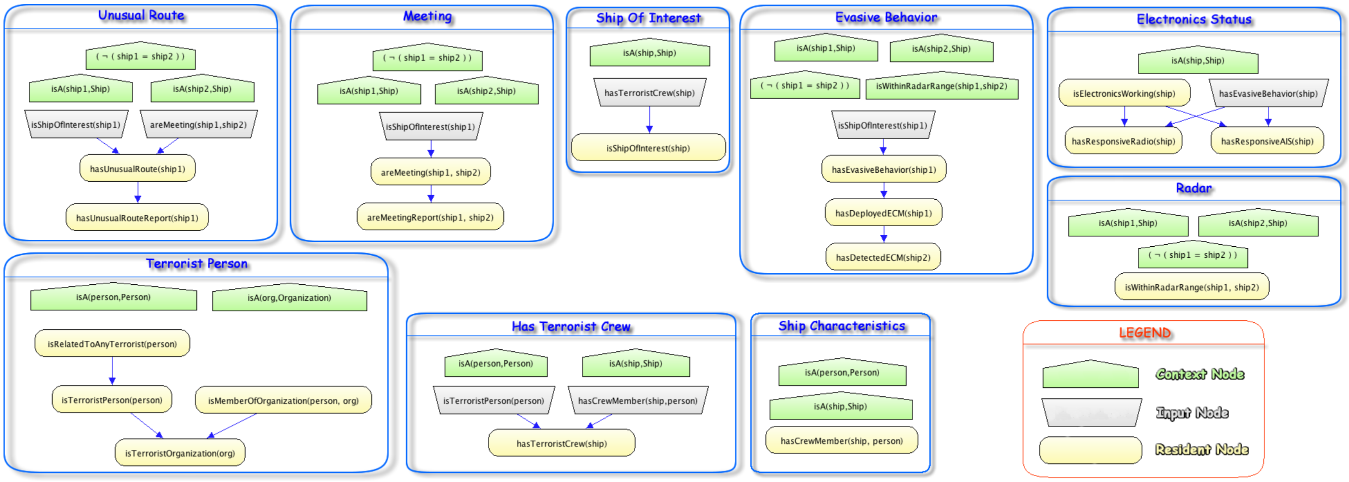

Figure 1. Probabilistic Ontology for Identifying Ship-of-Interest

B. *A PR-OWL Ontology for the Maritime Domain*

As an example of a PR-OWL ontology, Figure 1 shows a simple probabilistic ontology developed as part of the PROGNOS (Probabilistic OntoloGies for Net-centric Operation Systems) project [6]. The ontology is designed for the problem of identifying whether a vessel is a ship of interest. The model is designed to answer the following queries using the following evidence:

Overall Goal*: Identify whether a ship is a ship of interest, i.e. if the ship seems to be suspicious in any way.*

1. **Query**: Does the ship have a terrorist crewmember?
   a. **Evidence**: Verify whether a crewmember is related to any terrorist;
   b. **Evidence**: Verify whether a crewmember is associated with any terrorist organization.
2. **Query**: Is the ship using an unusual route?
   a. **Evidence**: Verify whether there is a direct report that the ship is using an unusual route;
   b. **Evidence**: Verify whether there is a report that the ship is meeting some other ship for no apparent reason.
3. **Query**: Does the ship seem to exhibit evasive behavior?
   a. **Evidence**: Verify whether an electronic countermeasure (ECM) was identified by a navy ship;
   b. **Evidence**: Verify whether the ship has a responsive radar and automatic identification system (AIS).

Each of the nine MFrags of Figure 1 addresses a modular component of the knowledge needed to address the above queries. Specifically, probabilistic knowledge about hypotheses related to the identification of a terrorist crewmember is represented in the *HasTerroristCrew*, *TerroristPerson*, and *ShipCharacteristics* MFrags. Knowledge about unusual routes is represented in the *UnusualRoute* and *Meeting* MFrags. Finally, knowledge about hypotheses related to evasive behavior is represented in the *EvasiveBehavior*, *Eletronics-Status*, and *Radar* MFrags.

A detailed explanation of this model can be found in [6]. The model was expanded and extended iteratively as described in [7] to address additional queries and evidence.

C. *An Open Source Tool for Probabilistic Ontologies*

The MFrags shown in Figure 1 are screenshots from the UnBBayes-MEBN [8], an open source, plug-in-based Java application for building and reasoning with probabilistic ontologies based on the PR-OWL/MEBN framework.[1] It features a graphical user interface (GUI), an application programming interface (API) for saving and loading PR-OWL ontologies, reasoning algorithms for processing queries, and plugin support for extensions.

D. *Queries*

Queries are processed in UnBBayes-MEBN using an implementation of the situation-specific Bayesian network (SSBN) construction algorithm described in [4]. Figure 2 shows an SSBN built using the implemented algorithm. We applied an exact inference algorithm on small-scale problems to test the model and identify logical inconsistencies, differences in query results from those expected by subject-matter experts, and other flaws in the model. For larger scale problems, approximate inference algorithms are employed to mitigate scalability issues. We also implemented hypothesis management methods [9] to control the complexity of the constructed networks while maintaining acceptable accuracy in results.

III. PR-OWL 2: Improving Compatibility with OWL

Ideally, it should be possible to use PR-OWL to reason probabilistically about uncertain aspects of an ontology based on the information already available. That is, we would like to

[1]UnBBayes is available from http://unbbayes.sourceforge.net/

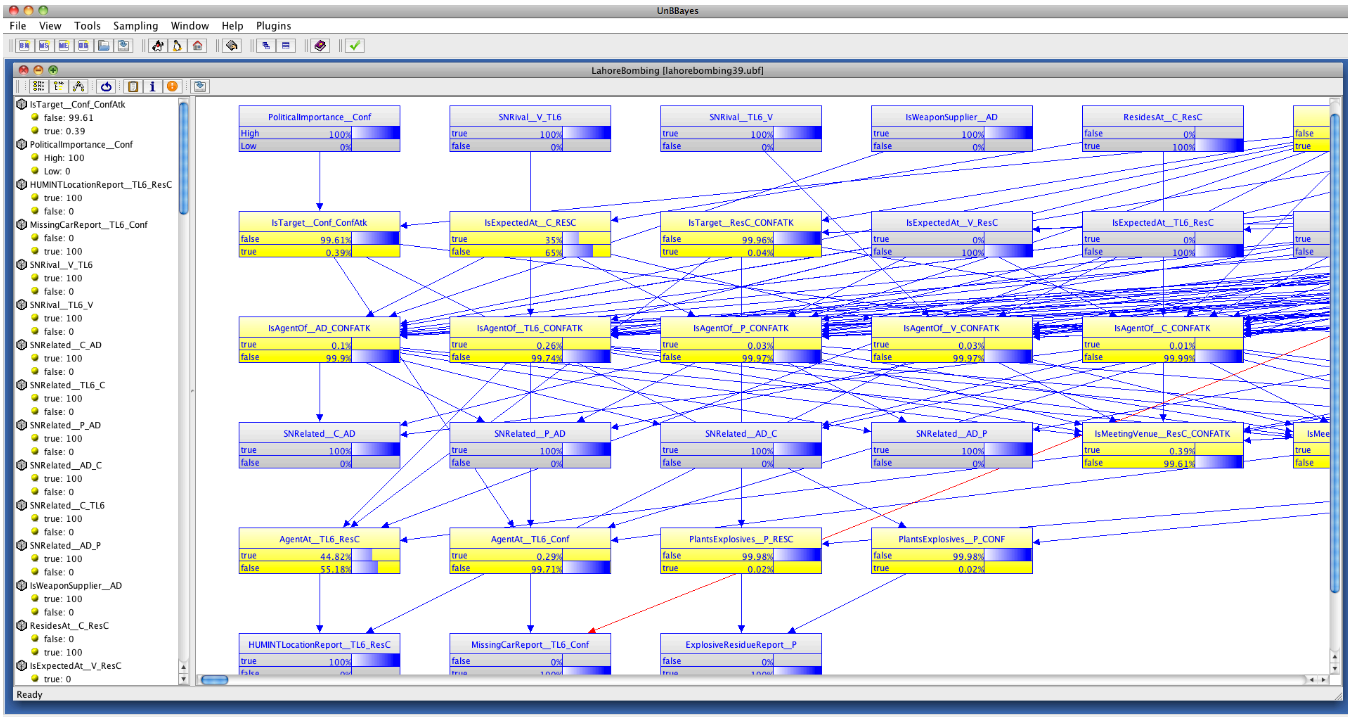

Figure 2. Situation-Specific Bayesian Network for Identifying Ship-of-Interest

be able to begin with an OWL ontology containing information about a domain, use PR-OWL to define uncertainty about attributes of and relationships among the entities, and apply a probabilistic reasoner to reason with available evidence. For example, we might begin with an OWL ontology containing classes for ships, routes, persons, and other entities mentioned in the MFrags of Figure1. We would then wish to use PR-OWL to define the probability distributions represented in the MFrags.

The difficulty with this idea is that PR-OWL 1 has no mapping between the random variables used in PR-OWL and the properties used in OWL. For example, suppose we have defined an OWL class Ship with property `isShipOf-Interest`, intended to represent whether a ship is a ship-of-interest. We might want to use the PR-OWL random variable `isShipOfInterest(ship)` to define the uncertainty associated with this property. We might use the *ShipOfInterest* MFrag of Figure 1 to specify its probability distribution. However, despite the syntactic similarity between the property name and the random variable name, PR-OWL 1 has no way to specify formally that the random variable `isShipOfInter-est(ship)` defines the uncertainty of the OWL property `isShipOfInterest`. Thus, even if we had information about whether a particular ship, say `Ship379`, is a ship-of-interest, we would not be able to instantiate the random variable `isShipOfInterest(ship)` for `Ship379`.

Poole et al. [10] point out the need to relate the random variables from probabilistic theories to the individuals, properties and classes of ontological languages like OWL. Poole et al. state, "We can reconcile these views by having properties of individuals correspond to random variables." This is the approach taken in PR-OWL 2.

The key to building the bridge that connects the deterministic ontology defined in OWL and its probabilistic extension defined in PR-OWL is to understand how to translate one to the other. On the one hand, given a concept defined in OWL, how should its uncertainty be defined in PR-OWL in a way that maintains its semantics defined in OWL? On the other hand, given a random variable defined in PR-OWL, how should it be represented in OWL in a way that respects its uncertainty already defined in PR-OWL?

PR-OWL 2 formalizes the relationship between OWL properties and PR-OWL random variables using the relation `definesUncertaintyOf` [3]. In our previous example, we would use the relation `definesUncertaintyOf` [3] to relate the OWL property `isShipOfInterest` to the PR-OWL 2 random variable `isShipOfInterest(ship)`. An additional complexity arises because MEBN can represent *n*-ary functions and predicates, whereas OWL has only binary properties. We must ensure that not only is the random variable linked to its associated OWL property by `defines-UncertaintyOf`, but also its arguments are linked to their respective OWL properties by either `isSubjectIn` or `isObjectIn`, depending on whether they refer to the domain or range of the OWL property, respectively. This feature is especially important when dealing with n-ary random variables, where each argument of the random variable will be associated with a different OWL property.

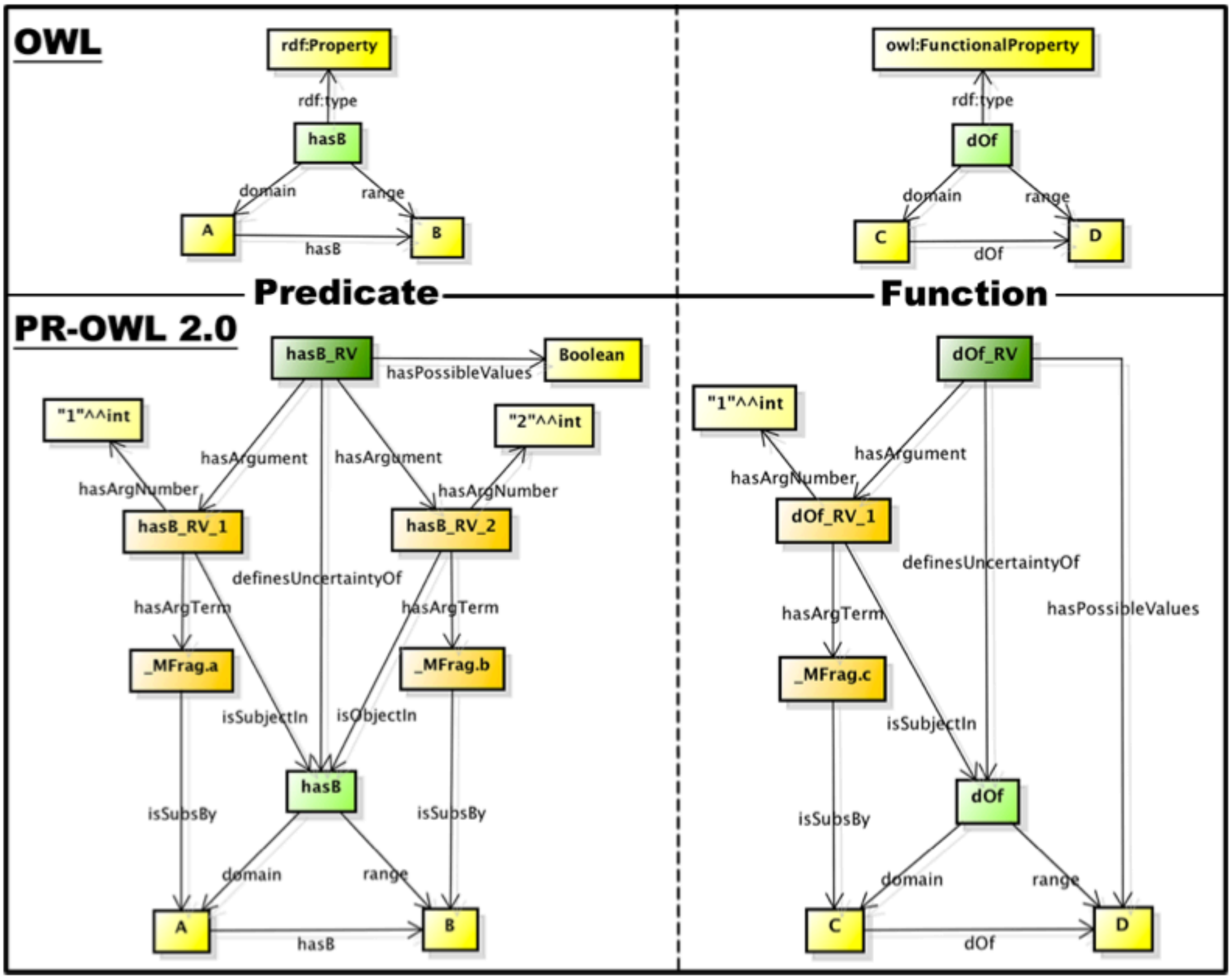

Figure 3. Mapping of PR-OWL Random Variables and OWL Properties

Figure 3 shows a schematic for the mapping between OWL properties and PR-OWL random variables. A full discussion of the formal mapping between OWL properties and PR-OWL random variables can be found in [3]. The mapping provides the basis for a formal definition of consistency between a PR-OWL probabilistic ontology and an OWL ontology, in which rules in the OWL ontology correspond to probability one assertions in the PR-OWL ontology. A formal notion of consistency can lead to development of consistency checking algorithms.

Another major difference between PR-OWL 1 and PR-OWL 2 is that the separate definition of entity in PR-OWL is replaced by OWL's built-in notion of classes and data types. That is, a PR-OWL entity is now identified with either a class or a data type in OWL. Moreover, since OWL supports multiple inheritance, so does PR-OWL 2. Thus, all the control over the type definition and type hierarchy in PR-OWL is delegated to OWL.

In PR-OWL 2, therefore, the possible values or outcomes of a random variable are instances of classes and data types. When specifying that a random variable will have individuals of a class as its possible outcomes, it is reasonable to assume that all known individuals of that class form a set of collectively exhaustive outcomes. However, the assumptions about individuals in OWL are not enough to guarantee these individuals are mutually exclusive. More specifically, although OWL provides a way to express unique names, it also allows two different names to point to the same object in the real world. To address this issue, PR-OWL 2 follows the MEBN and PR-OWL 1 convention, and assumes that every individual has a unique ID associated to it.

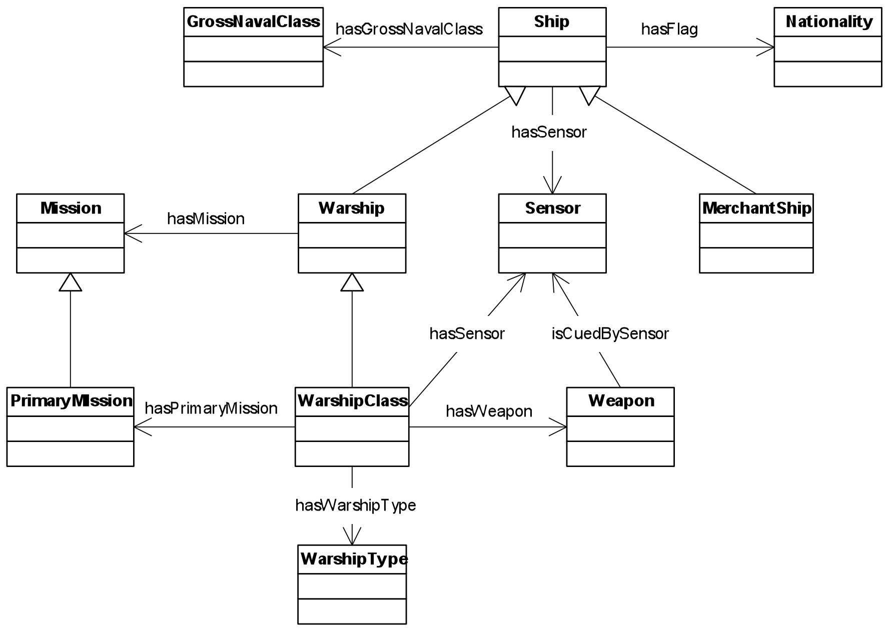

Figure 5. Entity-Relationship Diagram for Maritime Ship Ontology

We note that there are certain aspects of the full PR-OWL semantics that are not fully captured in OWL-DL, and therefore cannot be handled by OWL-DL reasoners, but are expected to be respected by PR-OWL reasoners. In particular, to specify the restriction that a random variable defines the uncertainty of a property would require OWL Full. For this reason, the restriction is not explicitly represented in PR-OWL, but it is expected to be enforced by a PR-OWL probabilistic reasoner. This enables consistency checking of the deterministic part of a PR-OWL ontology using a DL reasoner.

IV. PR-OWL 2 CASE STUDY

The following case study demonstrates the application of probability to an existing ontology to represent uncertainty in knowledge about instance attributes. In this case, an existing ontology of Western European warships identifies the major characteristics of each combatant class through the attributes of size, sensors, weapons, missions, and nationality. Figure 5 shows an entity-relationship diagram for the ontology. The decision maker is trying to determine the warship class of a contact about which he has limited information. By adding probability to the existing ontology, we can identify the most likely class of ship he is encountering when provided only partial or uncertain information. The model is designed to answer the following query using the following evidence:

Overall Goal*: Given uncertain or absent attribute information about a specific ship, what is the most likely European warship class that satisfies these attributes?*

1. **Query**: What is the type of warship?
    a. **Evidence**: Identify the size of the ship;
    b. **Evidence**: Confirm the ship is a warship;
    c. **Evidence**: Identify the primary mission of the ship based on its weapons and sensors.
2. **Query**: What nation has flagged the ship?
    a. **Evidence**: Identify the nation under which the ship is registered.

The entity-relationship diagram of Figure 5 presents a simplified design of the Military Ship Ontology illustrating the primary attributes used to answer these queries. The decision maker desires to know the class of warship that he faces. A class of ships has a consistent hull design and a standardized suite of weapons and sensors. These weapons and sensors work in concert to provide synergy in executing the primary mission of each type of ship. By combining a ship type with the nation that operates it, a logical prediction of warship class may be obtained.

International law of the sea requires that each merchant ship is registered and sails under a single nation for the purpose of regulation, certification, and pollution control. That process is known as flagging, and an individual ship is flagged by a nation. It is not required that a ship is flagged under the same nation as its owner; a “flag of convenience” allows a ship to be operated under an alternate nation to reduce operating costs and regulations. However, warships are always flagged under the nation of ownership.

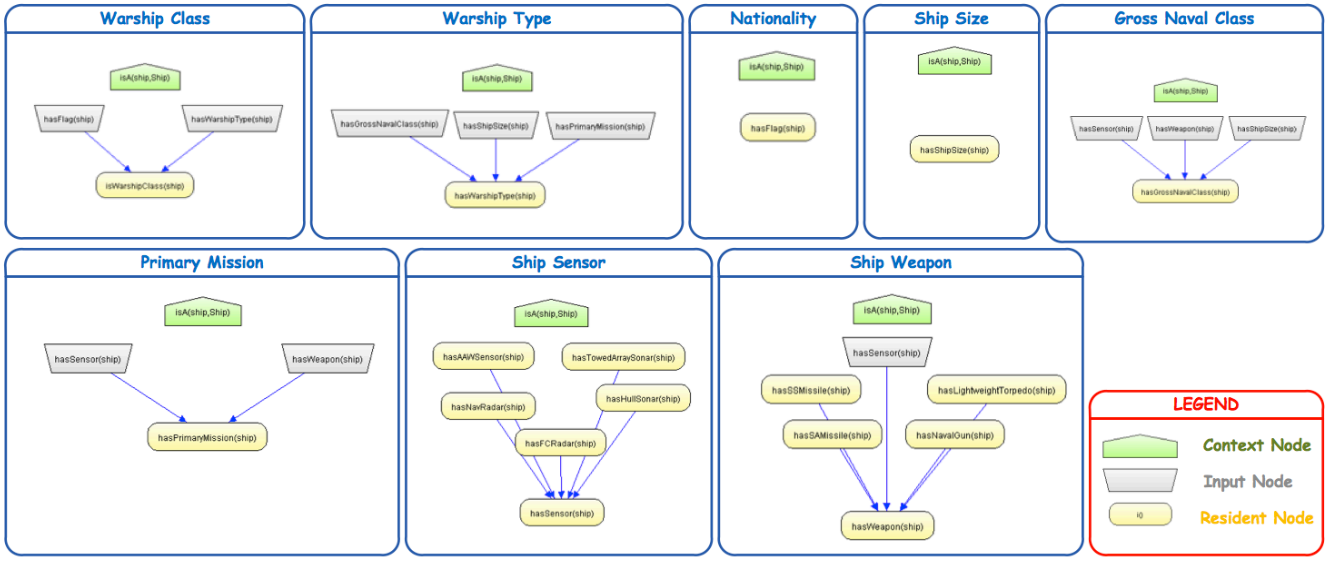

Figure 6. Military Ship Probabilistic Ontology

The Gross Naval Class is a naval schema that delineates warships from merchant ships, and is mutually exclusive. Through identification of weapon and sensor attributes, as well as overall ship size, a Gross Naval Class estimate may be made for the unknown ship. While it can be assumed that all ships have a radar sensor, only military ships have sensors associated with weapons systems. The presence of a weapon system, or a weapon-associated sensor, provides reasonable evidence that a ship is a warship.

Warships are of different types based on their primary mission. Most ships have multiple mission capabilities, but for this ontology we assume the following primary mission areas by ship type:

Anti-Air Warfare (AAW):
- Aircraft Carrier (CV, CVN)
- Cruiser (CG)
- Guided Missile Destroyer (DDG)
- Guided Missile Frigate (FFG)

Anti-Surface Warfare (ASuW):
- Destroyer (DD)

Anti-Submarine Warfare (ASW):
- Frigate (FF)

By observing the combination of weapons and sensors, it is possible to infer the most likely mission area. This, combined with an estimate of ship size, provides an indication of the type of warship.

At this point an MTheory is created to determine `hasWarshipClass(ship)` in the *WarshipClass* MFrag for some unknown ship. The eight MFrags associated with this determination are shown in Figure 6. Inputs to `hasWarshipClass` RV are the RVs from the *WarshipType* and *Nationality* MFrags, representing the concepts introduced above with the RVs `hasWarshipType(ship)` and `hasFlag(ship)`. The *WarshipType* MFrag may be further decomposed into the *ShipSize*, *GrossNavalClass*, and *PrimaryMission* MFrags. The *GrossNavalClass* MFrag is influenced by both the *ShipSize* and *ShipSensor* MFrags through the `hasShipSize(ship)` and `hasSensor(ship)` RVs, while the *PrimaryMission* MFrag is influenced by the *ShipSensor* and *ShipWeapon* MFrags with `hasSensor(ship)` and `hasWeapon(ship)` RVs. With the MTheory complete as shown in Figure 6, the Local Probability Distribution (LPD) must be populated.

Prior probabilities for the `hasFlag` RV were obtained from an estimate of merchant ship registrations available through open source information. Similarly, `hasShipSize` represents a finite and exhaustible set of ship lengths (`LengthLess150m, Length150to-100m, LengthGreater200m`) into which each `ship` is categorized. Prior probability estimates were again obtained via open source literature. Priors for `hasSensor` and `hasWeapon` were obtained through subject-matter-expert review of open source literature and represent the proportion of warships with each of the types of sensors. LPDs for the *GrossNavalClass* and *PrimaryMission* MFrags require conditional statements about relationships from the input nodes shown in Figure 6. A detailed description of these relationships is described in a forthcoming paper.

Queries to the Military Ship Probabilistic Ontology are processed in UnBBayes-MEBN using an implementation of the situation-specific Bayesian network (SSBN) construction algorithm. Instances of unknown ships and representative evidence are entered via the OWL ontology through the UnbBayes GUI to reflect partial or uncertain information

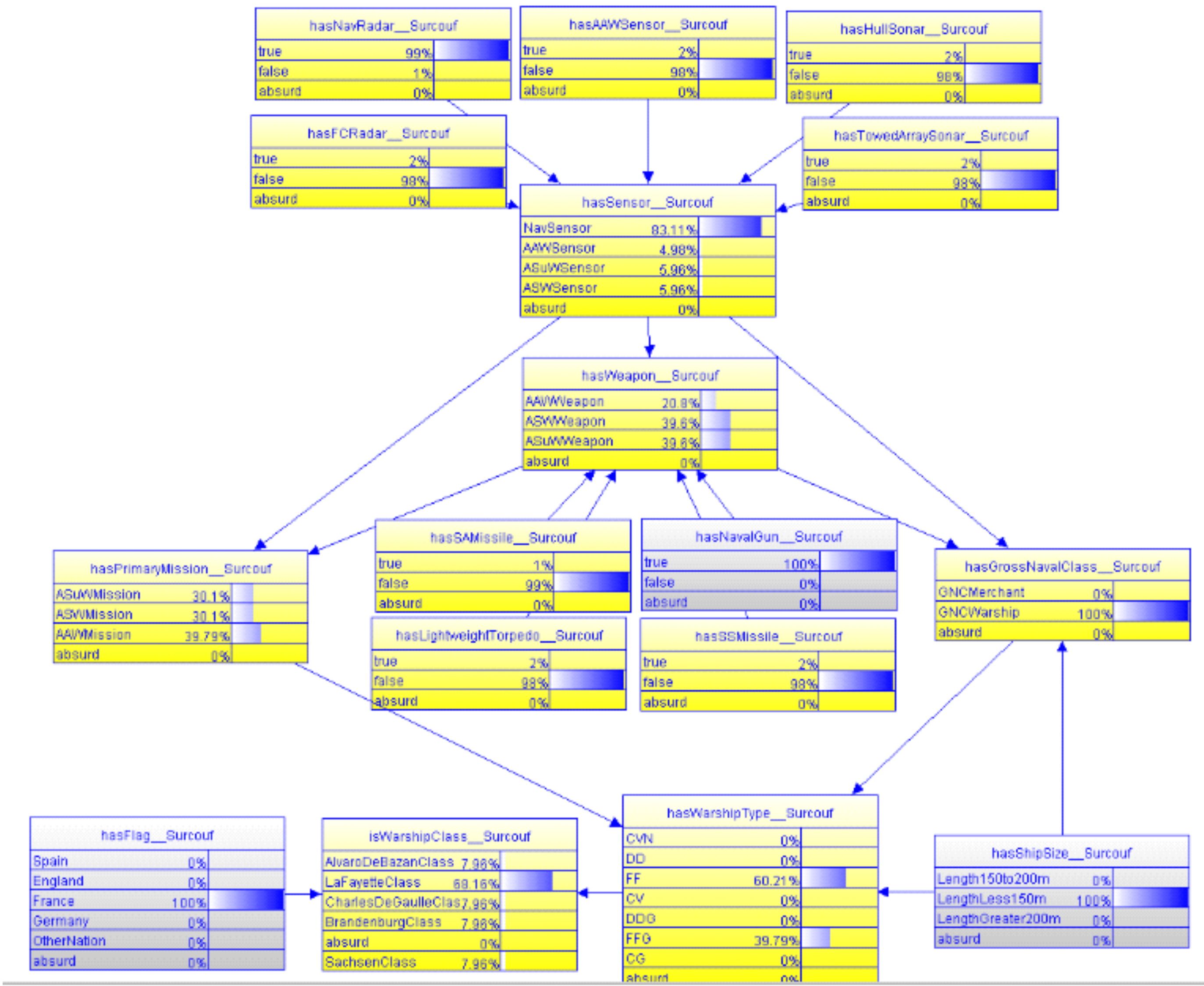

Figure 7 Situation-Specific Bayesian Network Military Ship Classification

about ship attributes. These are checked against known characteristics provided by subject-matter experts.

For example, suppose the following evidence is obtained about a ship of interest:

- `UID: Surcouf`
- `hasNavalGun(Surcouf): True`
- `hasFlag(Surcouf): France`
- `hasShipSize(Surcouf): <150m`

Executing a query of the `isWarshipClass` node produces the SSBN found in Figure 7. In this case, there is a 68% chance that *Surcouf* is a member of the French LaFayette Class of frigates, which is the correct classification.

As discussed in Section III, our goal is to begin with an OWL ontology containing information about a domain, use PR-OWL to define uncertainty about attributes of and relationships among the entities, and apply a probabilistic reasoner to reason with available evidence. Using the formalized construct introduced in PROWL-2, we map each of the RVs in the MFrags of the probabilistic ontology to the existing OWL property in the original ontology. This is accomplished through the probabilistic ontology building sequence executed on the UnbBayes software. For example, the `WarshipType` class in OWL has an object property of `hasPrimaryMission`. This object property is mapped to the `hasPrimaryMission(ship)` RV of the *PrimaryMission* MFrag. Mappings produced for each RV and its associated property in OWL allow us to use PR-OWL to reason probabilistically about uncertain aspects of an existing ontology based on the information already available.

V. Conclusion

Combining uncertainty reasoning with semantic technology is necessary for robust, interoperable, net-centric fusion and decision support systems. The probabilistic ontology language PR-OWL provides a way to represent and reason with probabilistic ontologies. PR-OWL 2 improves compatibility with OWL in several important respects. Through a case study, this paper describes the construction of a probabilistic ontology obtained by enhancing an existing OWL ontology with probability information.

References

[1] P. C. G. Costa, *Bayesian semantics for the semantic web*, PhD dissertation, Fairfax, VA, George Mason University (Jul. 2005).

[2] L. Predoiu, H. Stuckenschmidt, Probabilistic extensions of semantic web languages - a survey, in: *The Semantic Web for Knowledge and Data Management: Technologies and Practices*, Idea Group Inc, 2008.

[3] R. N. Carvalho, Probabilistic ontology: Representation and modeling methodology, PhD dissertation, Fairfax, VA, George Mason University (Jun. 2011).

[4] K. B. Laskey, MEBN: a language for First-Order bayesian knowledge bases, *Artificial Intelligence* 172 (2-3) (2008) 140–178.

[5] B. Milch, S. Russell, First-Order probabilistic languages: Into the unknown, in: Inductive Logic Programming, Vol. 4455 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2007, pp. 10–24.

[6] R. N. Carvalho, P. C. G. Costa, K. B. Laskey, and K. Chang, “PROGNOS: predictive situational awareness with probabilistic ontologies,” in *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.

[7] Carvalho, R.N., Haberlin, R., Costa, P., Laskey, K.B. and Chang, K.C., Modeling a Probabilistic Ontology for Maritime Domain Awareness. *Proceedings of the Fourteenth International Conference on Information Fusion*, July 2011.

[8] Matsumoto, S., Carvalho, R., Costa, P., Laskey, K.B., Santos, L.L. and Ladeira, M. There’s No More Need to be a Night OWL: on the PR-OWL for a MEBN Tool Before Nightfall. in *Introduction to the Semantic Web: Concepts, Technologies and Applications*, G. Fung, Ed. iConcept Press, 2011.

[9] R. Haberlin, P. C. G. da Costa, K. B. Laskey, Hypothesis management in support of inferential reasoning, in: Proceedings of the Fifteenth Inter- national Command and Control Research and Technology Symposium, Santa Monica, CA, USA, 2010.

[10] D. Poole, C. Smyth, R. Sharma, Semantic science: Ontologies, data and probabilistic theories, in: Uncertainty Reasoning for the Semantic Web I, Vol. 5327 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2008, pp. 26–40.