An Ontology for Hypothesis Management in the Maritime Domain

Richard J. Haberlin
Paulo Cesar G da Costa
Kathryn B. Laskey

June 22, 2011

Agenda

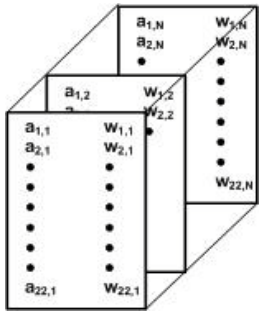
- Hypothesis Framework
- Maritime Domain Ontology
- Hypothesis Management Engine
- Simulation
- Results

Hypothesis Framework

Hypothesis Collection Framework

Hypothesis Framework

- Hypothesis Vector:
  - Arriving data stored as collection of 22 contextually relevant attributes
- Weight Vector:
  - Credibility assignment of each attribute in Hypothesis Vector
  - Derived from Source Pedigree of reporting unit



Hypothesis Knowledge Base

- Instantiated for each hypothesis in Hypothesis Knowledge Base
- Framework knowledge structure captures content and strength
  - Hypothesis vector describes a specific instantiation of a possible scenario
  - Weight vector allows us to update/compare its credibility

Query Hypothesis

- A domain-specific inquiry is posed by the system operator
  - Initiates the inferential reasoning process
  - Used to search for candidate hypotheses
  - Captured as an *m-tuple*
  - Compared with the stored metadata

- Associated *m x 1* Priority Vector
  - System operator prioritization of attribute fields
  - Used by HMM to retrieve and prioritize hypotheses from HKB

Maritime Domain Example

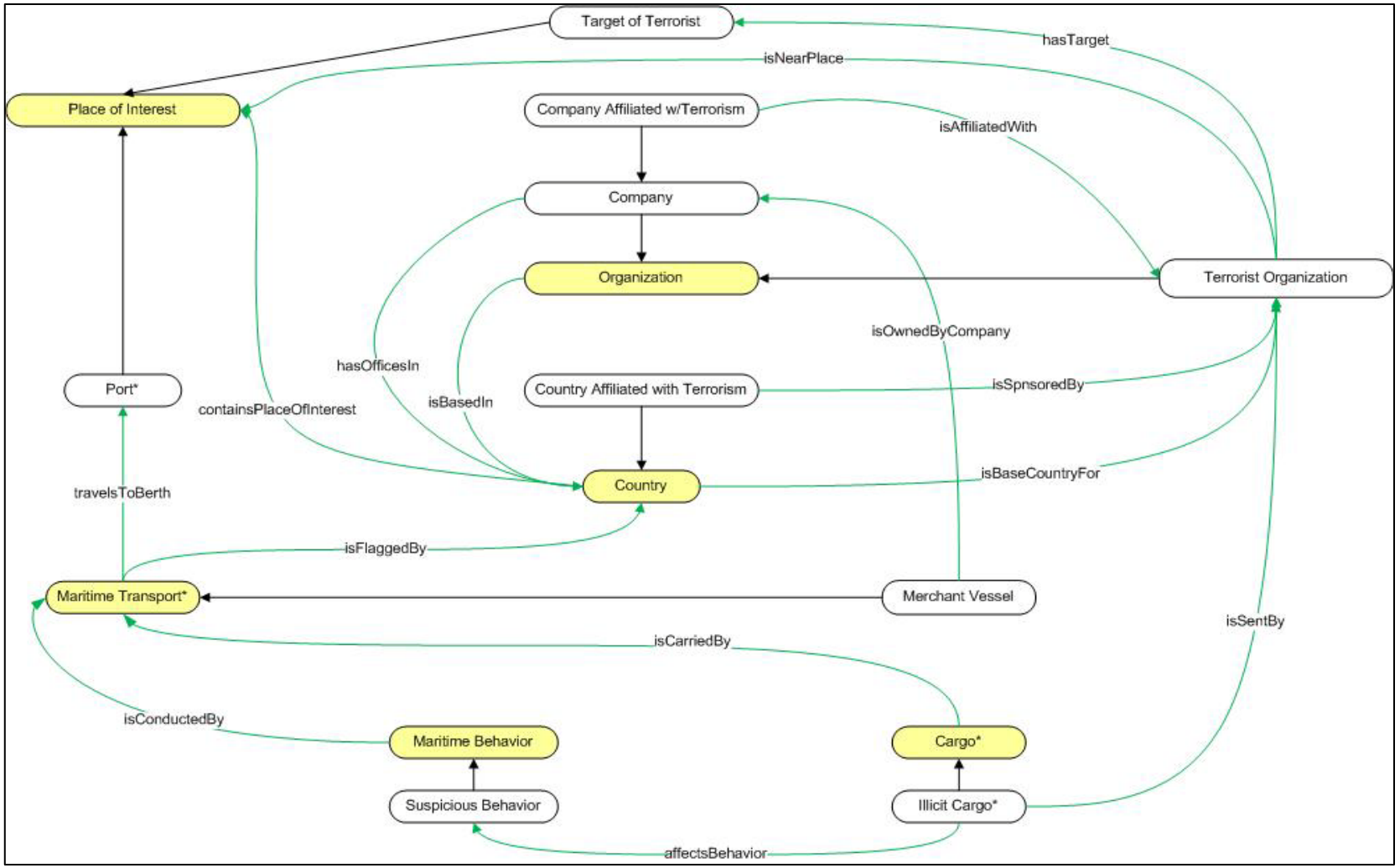
- Environment:
  - Mediterranean Sea
  - Atlantic Ocean
  - East Coast of North America
- Organization: Islamic Jihad Group
- Base of Operations: Izmir, Turkey
- Plan: smuggle radiological material
  - Arriving in Baltimore, Maryland
  - Bulk cargo vessel *Mustafa Kamal*
  - Build radiological dispersal devices

Maritime Domain Ontology

Ontology Background

- Merchant ships are a feasible means to smuggle illicit goods and personnel between nations
  - Multinational company ownership
  - Transit between coastal nations (global exposure)
  - Multinational and transient crews
- Ontology
  - Defines a common vocabulary for describing entities and relationships within a specific domain
  - Shares a common understanding of the structure of information
  - Ability to reuse the domain knowledge and structure for subsequent operation
- Maritime Domain Ontology
  - Captures the maritime domain to assist in the situational awareness problem
  - Describes the platforms and states of maritime vessels
  - Ports of departure and arrival are limited to those in the Mediterranean Sea and Atlantic Ocean
  - Cargo is shipped from departure ports in the Mediterranean to ports in North America
  - HKB constructed of class instantiations with defined attribute values and additional relationships

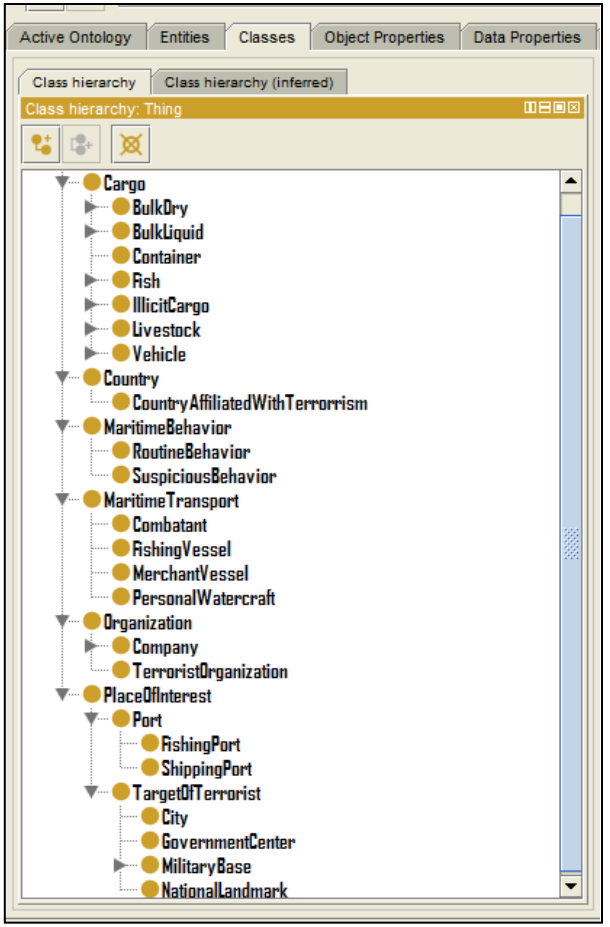
Maritime Domain Ontology



framework | **maritime domain ontology**| hypothesis management engine | simulation| results

Ontology Classes

- Breadth: 74 Classes (47 cargo types); 6 Super Classes
  - Cargo
    - May be legal or illicit
  - Country
    - Contains Places of Interest & corporate offices
    - Flag maritime transports
    - Base terrorist organizations
  - Maritime Behavior
    - Mariners maximize profit and minimize risk
    - Deviations are suspicious
  - Maritime Transport
    - Merchant vessels, fishing vessels, combatants
  - Organization
    - Define social structure of individuals
    - Located in countries, may consist of or be affiliated with terrorists, and may own/operate maritime transports.
  - Place of Interest
    - Represent departure and destination points
    - May be targets for terrorist organizations



Ontology Implementation

- Captured in Protégé Version 4.1.0 (Build 213)
- Provides domain-specific ontology for inferential reasoning systems
  - Specifies domain using real-world data
  - Produces more realistic output
- Future builds
  - Smaller fishing (only) ports
  - More rigorous evaluation of relationships between terrorist organizations and companies

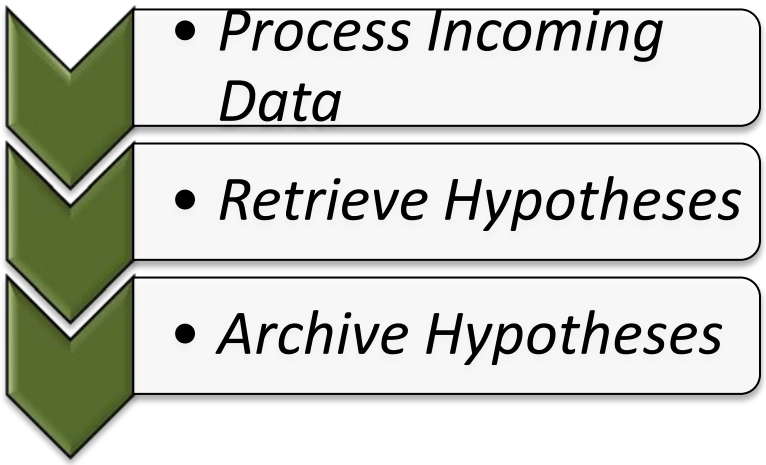
Unit	Number	Source
Cargo Classes	47	Subject-matter Expert Data
Country Individuals	78	CIA World Factbook 2010 [2]
Maritime Behavior Properties	41	Subject-matter Expert Data
Maritime Transport Classes	4	Subject-matter Expert Data
Terrorist Org. Individuals	46	U.S. Dept. of State List [5]
Company Individuals	121	Directory of Top Int'l Maritime Shipping Lines & Ship Owners [9]
Port Individuals	73	CIA World Factbook [2]
City Individuals	37	Subject-matter Expert Data
Target Individuals	51	Subject-matter Expert Data

Domain Relationship Implications

- MDO contains many relationships between classes
  - Some imply increased likelihood of association with terrorism
  - Some imply a ship is of interest in maritime domain awareness
- Relationships indicating a suspicious relationship
  - Company Affiliated with Terrorism owning a Maritime Transport
  - Country Sponsoring Terrorist Organization flagging a Maritime Transport
  - Target of Terrorist is near Terrorist Organization
  - Port is near Terrorist Organization
  - Company has offices in Country Sponsoring Terrorist Organization

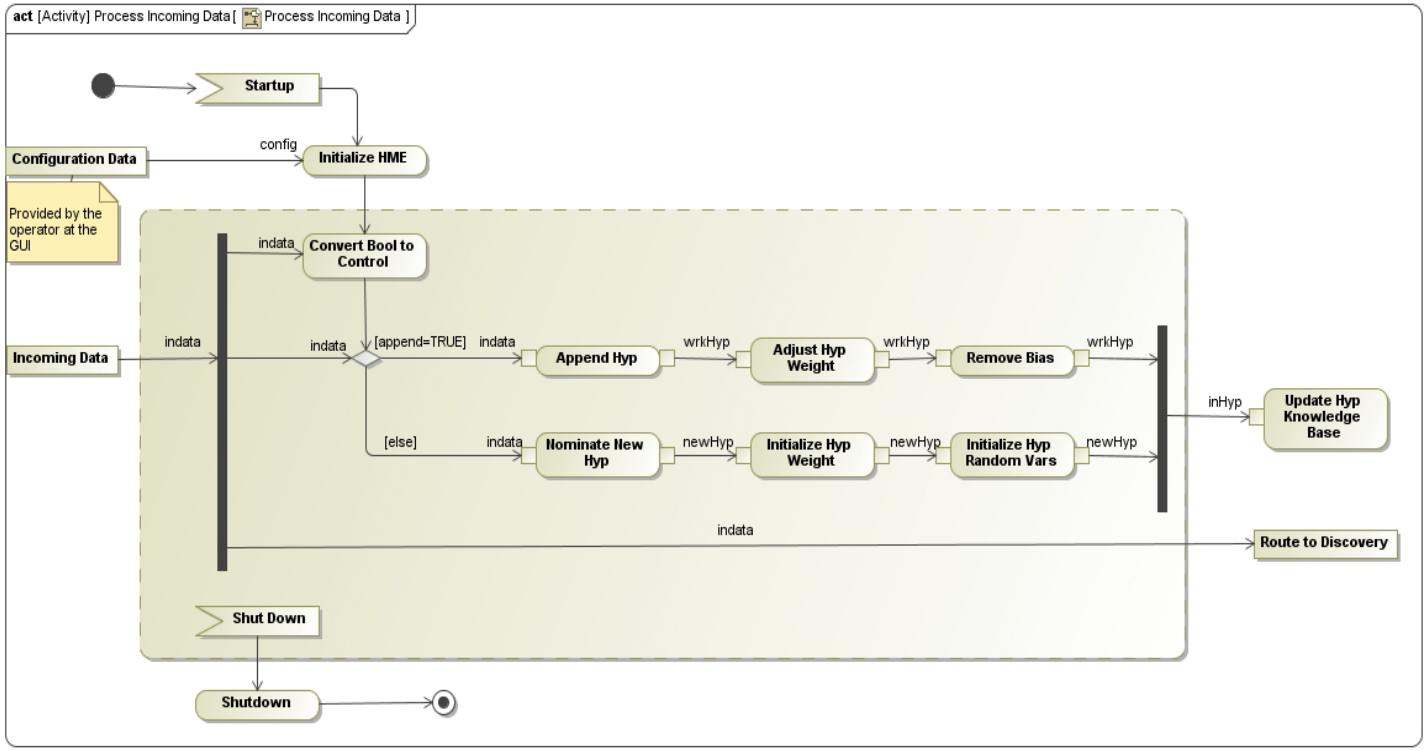
Hypothesis Management Engine

Hypothesis Management Engine



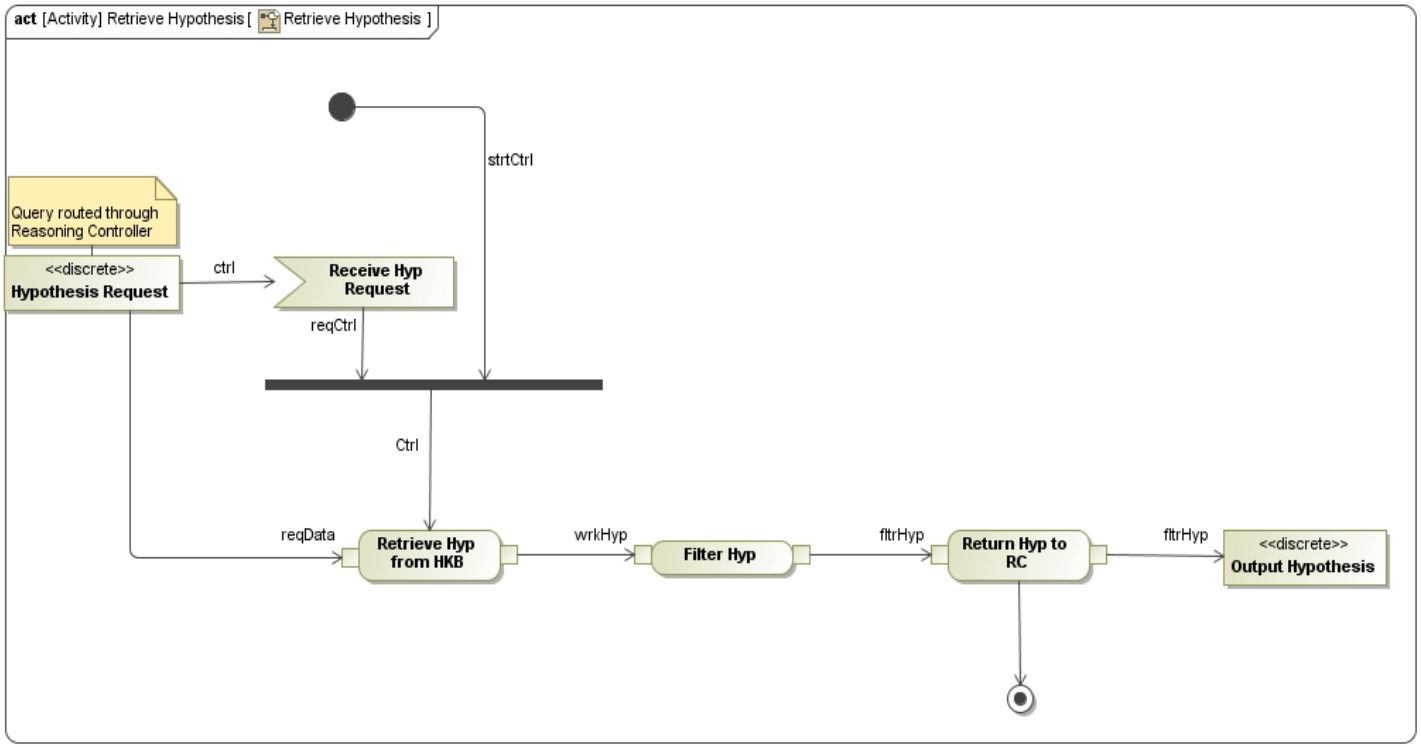
- Creates, updates, administrates, filters and routes hypotheses
- Coordinates with HKB for retrieval and storage of hypotheses
- Delivers set of contextually relevant hypotheses for inferential reasoning as a result of an operator query

Process Incoming Data Activity



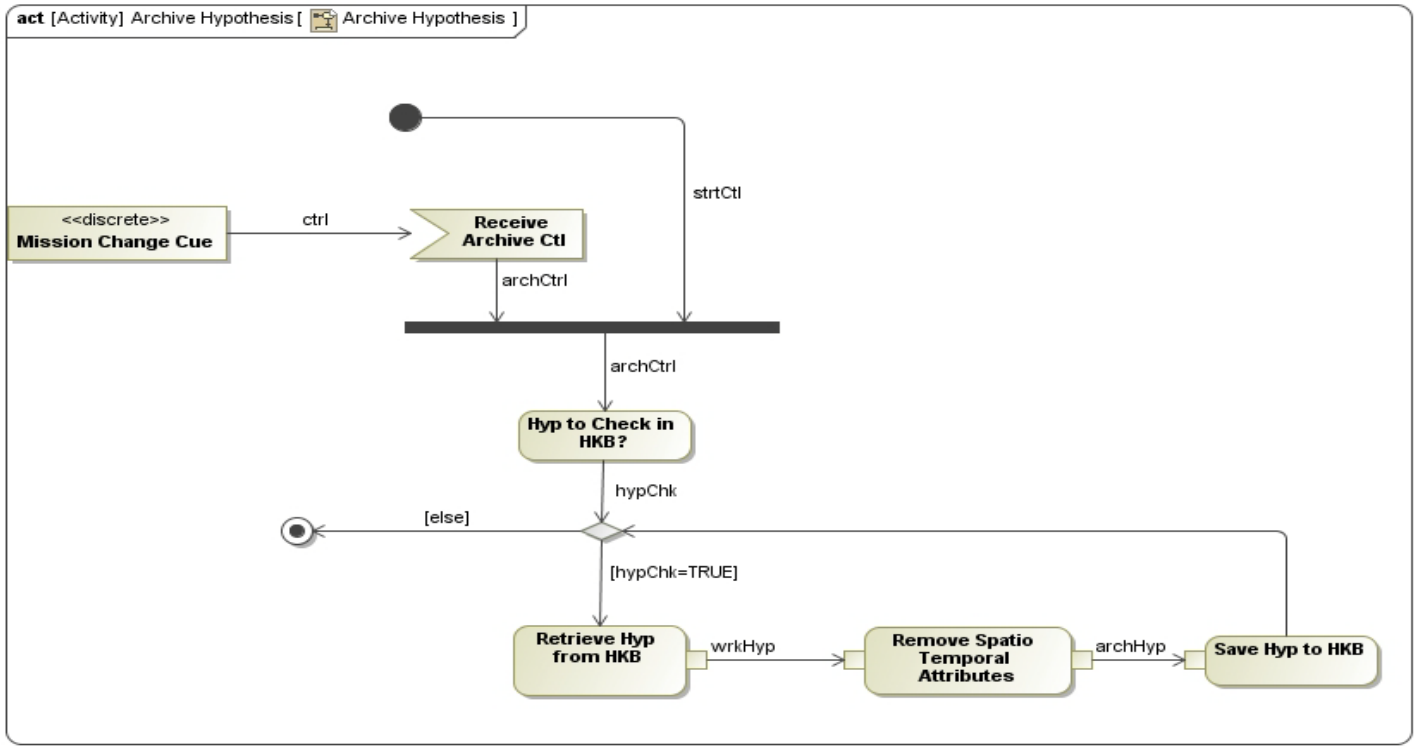
- Continuously creates and updates hypotheses from data
- Operations within the shaded interruptible region execute continually on incoming streaming data until shutdown

Retrieve Hypothesis Activity



- Reasoning Controller requests candidate hypotheses
- HME coordinates with the HKB for retrieval, filters and prunes the hypotheses within the context of the query, and forwards the filtered hypotheses

Archive Hypothesis

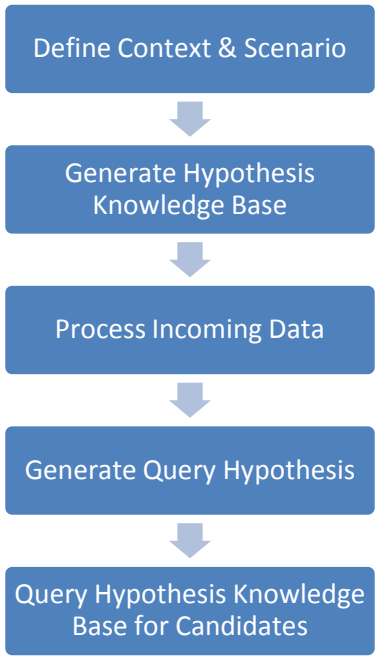


- Allows non-time sensitive attributes of hypotheses to be archived in the HKB in anticipation of building upon them upon return to the area of operations

Simulation

Simulation Methodology

- Provides an opportunity to observe HME in a synthetic environment
  - Scenario with HME running for a short period of time
  - Number of hypotheses (100) gathered in the HKB
- Receives a varying number of inputs that must be processed for inclusion and for possible relationships
- HKB searched for candidate matches to a randomly generated hypothesis query



Generating the Random HKB

- Randomly created a contextually accurate HKB of 100 entries
  - Classes, properties, and individuals from the MDO
  - Each entry has 30% probability for each of 22 attribute fields
- Assumptions affecting HKB density
  - Size of the track pool determines the frequency with which the unit identification numbers are repeated
    - A track pool of 10,000 entries was allowed, making duplication unlikely.
  - Each unit captain has an identification number
    - Pool of 1000 captains makes duplication of a captain significantly more likely when updating incoming information.
  - The probability of inclusion of 30% discussed above determines if an attribute is to be included in a hypothesis
    - Greater inclusion would create a KB of more-dense hypotheses

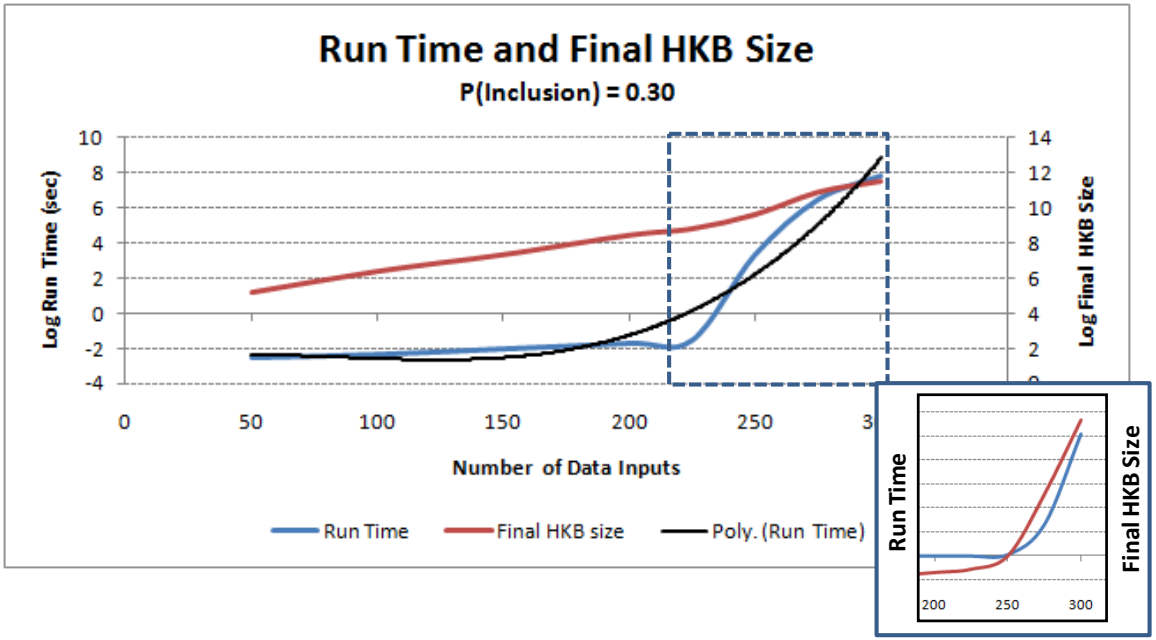
Generate Random Query Hypothesis

- Randomly created a query hypothesis for each run
  - Classes, properties, and individuals from the MDO
  - Each entry has 70% probability for each of 22 attribute fields
  - High likelihood that each attribute is of interest to operator
- Single query hypothesis created and executed against the randomly created, 100-entry HKB and its additional inputs using the *Retrieve Hypotheses* algorithm

```
Sort Query Hypothesis attributes by Priority Vector value
For each priority value above a threshold {
For each non-zero attribute {
        Check each HKB entry for matching value
        If values match, return hypothesis as candidate}}
```

Results

Results



- Natural log of p size shown for simulation run with P{Inclusion} = 0.30 and starting HKB size of 100 entries
- Linear profile of the HKB size denotes exponential growth rate ~ $O(N^2)$
- Exponential plot for natural log of run-time indicates a higher-level polynomial relationship for processor time represented by the black trend line

Summary

- Context-variable relationships drive the computational complexity of the HME
  - Reductions in overhead only realistically achieved by reducing the number of context variables for the *Process Incoming Data* activity
- Linear profile of the HKB size denotes exponential growth rate (O(N2))
  - Primarily caused by creation of context relationships evolving from multiple hypothesis possibilities
  - Volume of classes, properties and specified individuals also affect performance
  - Complexity driven primarily by number of context variables included
- Exponential plot for the natural log of run-time indicates a higher-level polynomial relationship for processor time

An Ontology for Hypothesis Management in the Maritime Domain

References

- [1] R. Haberlin, P. C. G. da Costa, and K. B. Laskey, "A Model-Based Systems Engineering Approach to Hypothesis Management," in Proceedings of the 3rd International Conference on Model Based Systems Engineering, 2010.
- [2] "CIA - The World Factbook," Central Intelligence Agency - The World Factbook, Jan-2011. [Online]. Available: https://www.cia.gov/library/publications/the-world-factbook/. [Accessed: 18-Jan-2011].
- [3] "DAML Ontology Library," DAML Ontology Library, Jan-2011. [Online]. Available: http://www.daml.org/ontologies/. [Accessed: 18-Jan-2011].
- [4] R. Haberlin, "Establishing a Source Pedigree Ontology Through Evidential Reasoning," Dec-2010.
- [5] "Foreign Terrorist Organizations," US Department of State Foreign Terrorist Organizations, Jan-2011. [Online]. Available: http://www.state.gov/s/ct/rls/other/des/123085.htm. [Accessed: 18-Jan-2011].
- [6] "Freebase," Freebase, Jan-2011. [Online]. Available: http://www.freebase.com/. [Accessed: 18-Jan-2011].
- [7] R. Haberlin, P. C. G. da Costa, and K. B. Laskey, "Hypothesis Management In Support of Inferential Reasoning," in Proceedings of the 15th International Command and Control Research Symposium (ICCRTS), 2010.
- [8] "Linked Data | Linked Data - Connect Distributed Data across the Web," Linked Data, Jan-2011. [Online]. Available: http://linkeddata.org/. [Accessed: 18-Jan-2011].
- [9] "Maritime Shipping Companies : Ship Owners," Directory of Top Int'l Maritime Shipping Lines & Ship Owners, Jan-2011. [Online]. Available: http://www.onelasvegas.com/wireless/ship_owners.html. [Accessed: 18-Jan-2011].
- [10] "ODP - Open Directory Project," DMOZ Open Directory Project, Jan-2011. [Online]. Available: http://www.dmoz.org/. [Accessed: 18-Jan-2011].
- [11] "Ontolingua Home Page," Ontolingua, Jan-2011. [Online]. Available: http://www.ksl.stanford.edu/software/ontolingua/. [Accessed: 18-Jan-2011].
- [12] J. Boyd, "Organic Design for Command and Control" [Electronic Version]. Unpublished lecture notes. Retrieved January, 2011, from: http://www.ausairpower.net/JRB/c&c.pdf.
- [13] P. C. G. da Costa, K. B. Laskey, and K. Chan, "PROGNOS: Applying Probabilistic Ontologies to Distributed Predictive Situation Assessment in Naval Operations," in Command and Control Research and Technology.
- [14] "RosettaNet Home," RosettaNet, Jan-2011. [Online]. Available: http://rosettanet.org/. [Accessed: 18-Jan-2011].
- [15] "The Protégé Ontology Editor and Knowledge Acquisition System," Protege, Jan-2011. [Online]. Available: http://protege.stanford.edu/. [Accessed: 19-Jan-2011].
- [16] "UNSPSC Homepage," United Nations Standard Products and Services Code, Jan-2011. [Online]. Available: http://unspsc.org/. [Accessed: 18-Jan-2011].