ISRCS 2008 Keynote Address (September 9, 2008)

On Threats from Intelligent Adversaries

Eugene Santos Jr.
Thayer School of Engineering
Dartmouth College
Eugene.Santos.Jr@dartmouth.edu

Outline

- Traditional approach to control system resilience(?)
  - Deficiencies of Traditional Approaches when Confronted by Intelligent Adversaries
- What is Intent?
- Modeling Intelligent Adversaries
- Examples of Intent Inferencing
- Conclusion and Discussion

Control System Qualities

- Many terms developed to describe desirable qualities of a control system related to resilience
  - Stability
    - BIBO, Lyapunov, etc.
  - Robustness
    - Control of unknown plants with unknown dynamics subject to unknown disturbances (Chandrasekharan '96)
  - Survivability
    - The aggregate loss from a disturbance is minimized and judged acceptable under defined adverse circumstances (Ellison et al '97)
  - Adaptability
    - Ability to change system parameters over time to deal with time-varying parameters or uncertainty in the system (Sastry '89)
  - Security
  - Resiliency

Resilience

- Three main factors to consider
  1. the ability to resist the effects of disturbances and keep system parameters within some safe range to whatever degree is possible
  2. the ability to gracefully fail in the face of disturbances (avoid catastrophic damage)
  3. the ability to easily recover after the disturbance has passed

Stability

- BIBO Stability (Proakis '96)
  - Given bounded inputs, ensure that outputs will also be bounded
- Lyapunov Stability (Lyapunov '66)
  - Ensure that the output remains close to a target value
- Root-Locus (Franklin '93)
  - System is stable if all roots of the transfer function are on the left hand side of the plane in the frequency domain

Robustness

- Kalman Filtering (Kalman '60)
  - Provide estimations of system state in the face of noise and incompleteness
- H-Infinity Control (Chandrasekharan '96)
  - Hankel norms are used to measure control system properties and minimize effects of perturbations

Dealing with Uncertainty

- Stochastic Control
  - Uncertain parameters are modeled as probability distributions
- Fuzzy Control
  - Useful for systems that are difficult to specify analytically

Techniques

- These techniques have had success at mitigating some factors that can compromise resiliency
  - Modeling errors
    - Uncertainty about system parameters and expected inputs
  - Noise in sensor inputs
- But they have their issues as well
  - Often expect bounded inputs
  - Often do not assume that nature will do its worst (Lainiotis '76)
  - They are largely *reactive*, i.e. try to fix the problem when we notice it. Need to be more *proactive*.

Intelligent Adversaries

- What if nature is not our only problem?
- Threats posed by intelligent adversaries have some key differences from natural threats
  - Nature does not always try to do its worst
    - Must assume that an adversary will
  - Intelligent adversaries may have inside information
    - Can target the most vulnerable aspects of a system
  - Uncertainty
    - High degree of uncertainty about how the adversary may attack
  - Incompleteness
    - Must be able to make decisions in the face of incompleteness

Increased Accessibility

- In the past control systems were often isolated and run with proprietary equipment
- Not anymore (Cárdenas et al '08)
  - Often accessible through the internet
  - Use open protocols and off the shelf components familiar to many
- Provides a familiar attack vector for intelligent adversaries

Recent Efforts

- Several recent efforts have focused on the susceptibility of control systems to cyber attacks from intelligent adversaries
  - Control Systems Security Program (CSSP) initiated by DHS
    - "coordinates activities to reduce the likelihood of success and severity of impact of a *cyber attack* against critical infrastructure control systems through risk-mitigation activities."
  - Control System Security Center at Idaho National Laboratory
    - "efforts to secure the computer-aided control systems that operate the nation's critical infrastructures"
  - Survivability and Recovery of Process Control Systems project at the I3P
    - "Process Control Systems are thus vulnerable to the numerous cyber threats—both malicious and inadvertent"
    - "researchers from eight leading academic institutions, federally-funded labs and non-profit organizations across the U.S. are engaged in an intense effort to increase PCS resiliency and strengthen the nation's critical infrastructures"

Why Adversarial Modeling?

- "The industry is largely unaware of the threat environment and adversary capabilities"
  - Sandia National Labs, *Common Vulnerabilities in Critical Infrastructure Control Systems*

Why Adversarial Modeling?

- Allows for *proactive* measures to be taken
- Many factors to consider to achieve this
  - What are the pertinent threats and adversaries?
  - Which targets are most attractive to these adversaries?
  - What tools do these adversaries have at their disposal and how effective are these tools against a given target?

Why Adversarial Modeling?

- Allows reasoning about threats
  - Deductive Reasoning: Given what we know about the adversary, what can we conclude?
    - What are the likely targets and methods of attack?
  - Abductive Reasoning: Reason from the observed evidence to the best hypothesis that explains the evidence
    - What are goals and beliefs of the adversary that best explain the types of attacks that are observed?
- This in turn provides guidance for the more traditional approaches to resilience in control theory
  - The types of disturbances they should plan for

Outline

- Traditional approach to control system resilience
  - Deficiencies of Traditional Approaches when Confronted by Intelligent Adversaries
- What is Intent?
- Modeling Intelligent Adversaries
- Examples of Intent Inferencing
- Conclusion and Discussion

Behavior

- What: Actions, Activities, Decisions, Maneuvers, Operations, Performance, etc.
- How: Sequences of, Courses of Actions, Plans Executed, Steps Taken
- Who: Individuals, Machines, Groups, Organizations
- **Why:** Need to understand the basis for behavior – Intent
  - Also, dynamic, emergent, etc.

What is Intent?

- Intent inferencing, or user intent inferencing, involves deducing an entity's goals based on observations of that entity's actions (Geddes, 1986)
  - Deduction involves the construction of one or more behavioral models that have been optimized to the entity's behavior patterns
  - Data/knowledge representing observations of an entity, the entity's actions, or the entity's environment (collectively called observables) are collected and delivered to the model(s)
  - Models attempt to match observables against patterns of behavior and derive inferred intent from those patterns
- Userful for generation of advice, definition of future information requirements, proactive aiding, or a host of other benefits (Bell et al., 2002; Santos, 2003)

Approaches to Intent Inferencing

- **Plan-goal-graph (PGG)** – a network of plans and goals, where each high level goal is decomposed into a set of plans for achieving it, and the plans are decomposed into subgoals which in turn are decomposed into lower-level plans (Geddes, 1994)
  - Intent is finding the path from observables to a plan or goal

Approaches to Intent Inferencing

- Operator function model (OFM) – an expert system using a heterarchic-hierarchic network of finite-state automata, in which nodes represent entity's activities and arcs represent conditions that initiate/terminate certain activities (Bushman et al., 1993; Chu et al., 1995; Rubin et al., 1988)
  - Connect observed action to appropriate activity trees

Approaches to Intent Inferencing

- Generalized plan recognition (GPR) – recognize the entity's plan for carrying out the task, based on observations, an exhaustive set of discrete actions (a plan library), and constraints (Lesh et al., 1998; Carberry, 1988; Goodman and Litman, 1990)

- Others –BDI, GOMS, etc.

Intent Inferencing

- Approaches mentioned so far mainly for static plan recognition
- Can be ambiguous on who/what being modeled
- Only secondary focus on uncertainty
- Team interaction not well-defined
- Explore strengths and weaknesses of approaches and applicability to our problem domain

Why for Intent?

- It's about modeling others. (So we can better understand them.)
  - The User
    - Machine is working with the User so the Machine wants to know intent of the User
  - The Machine
    - User is working with the Machine so the User wants to know the intent of the Machine – UA/GVs
  - The Organization
    - …

What is "Other's" Intent?

- What's the context of an action they took?
- What is the rationale behind the action?
- What are the causes and effects of their intended goal?
- What is the motivation behind a specific behaviour?
- What will happen next?
- Why did this behaviour occur?
- What do they believe?

Accounting for Human Factors in Capturing Intent

- Assymetric – they are not like us/me; I/we do not think like them
  - E.g., "What is rational" is not the same between different individuals or groups especially with different backgrounds.
- Differences in decision-making and behavior come from differences in background
  - Experience
  - Social
  - Cultural
  - Economic
  - Political
  - Psychological
  - Training

"Other's" Intent

- Intent is not just the plan or course of action of the other entity
- Not just "The enemy commander *intends* to launch his SAMs", "The organization *intends* to undertake a suicide bombing", "The surgical team intends to remove the brain tumor with clear margins", or "The corporation *intends* to 'go green' ", but also why??

- Intent is the highest-level goal(s) they are pursuing + the support for that goal + the plan to achieve it

- Need intent to understand and predict other's behavior
- Must model them based on their **perceptions** of the world
- (Santos 2003; Santos & Zhao 2006)

Intent – What can you do with it?

- Forecast the future: actions, reactions, behaviours, needs, etc.
- Explain the present: causes, motivations, goals, tasks, etc.
- Understand the past: beliefs, axioms, history, experience, etc.

- Knowing intent is crucial to communications and collaborations:
  - Human-human, human-machine, machine-machine, human-organization, organization-organization
  - Complex organizations consist of multiple humans and machines
- Inferred intent knowledge can also help focus and prune search space, bound optimization, guide scheduling, and better allocate resources.
- (Bell et al. 2002; Santos 2003; Santos & Zhao 2006)

Outline

- Traditional approach to control system resilience
  - Deficiencies of Traditional Approaches when Confronted by Intelligent Adversaries
- What is Intent?
- Modeling Intelligent Adversaries
- Examples of Intent Inferencing
- Conclusion and Discussion

Challenges

- Each individual or group is a unique entity
- Human factors are difficult to capture accurately and/or completely
- Uncertainty associated with the impacts of human factors on decision-making process is inherent
- Intent and behavior evolves over time
- Underlying problem of abductive inference
  - Intuition – explain the source of the observations
- **"If you build it, does it really work?"**

Modeling and Perception

- Our Approach: Model of entity based from entity's perception or point of view
  - How does the entity view the world?
  - What can the entity observe about us and others?
  - Explanation of entity behavior grounded in terms of the entity's world-view
    - Avoids accidentally imposing our beliefs on the entity
- Observables and evidence passed to the entity model is based on the above questions
  - Obviously, the entity does not see everything
  - Allows for modeling of deception (and self-deception?)

An Intent Modeling Approach

- Incorporate human factors
- Intent driven
- Model the decision making process based on how "other's" views the world
- Build network fragments for each piece of information / knowledge, and merge them together for reasoning
- Based on Bayesian Knowledge Bases (BKBs) [Santos & Santos 97]

Intent Driven Approach

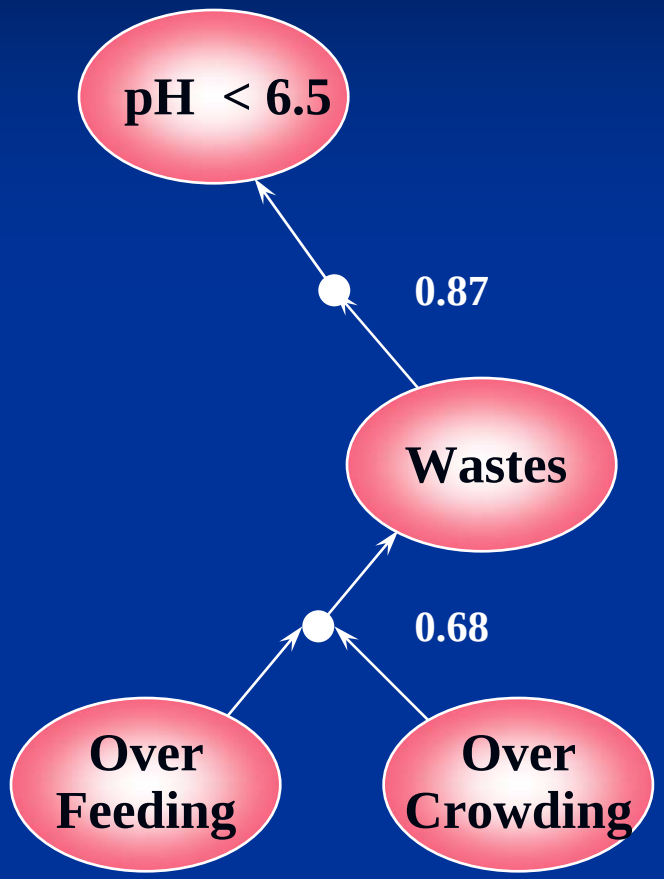
- **Model adversary through 3 formative components:**
  - **Goals/Foci:** A prioritized (by probability) list of short and long term goals representing entity intents, objectives or foci. The goal component captures ***what*** the entity is doing.
  - **Rationale Network:** A probabilistic network representing the influences of the entity's beliefs, both about themselves and others, on their goals and on high level actions associated with those goals. The rationale component infers ***why*** the entity is behaving in a certain fashion.
  - **Actions Network:** A probabilistic network representing the detailed relationships between entity goals and possible actions to realize those goals. The action component captures ***how*** an entity might act.
  - (Santos 2003; Santos & Negri 2004; Santos & Zhao 2006)

Bayesian Knowledge-Bases

- Simple method of knowledge representation
  - "if-then" rules with conditional probabilities
- Mathematically sound model
- Subsumes existing knowledge representations
  - Bayesian Networks [Pearl 1988; Pearl 2000]
- Handles incomplete and cyclic information
- Eases problems in acquisition, V & V
  - Automated correction, fine-tuning, and learning
- Bayesian Knowledge Fragments for aggregation
  - Ready (dis-)aggregation through knowledge fusion, scalability
  - Basis for quick reaction reasoning
  - Natural modularity to capture changing intentions, goals, and decision-making

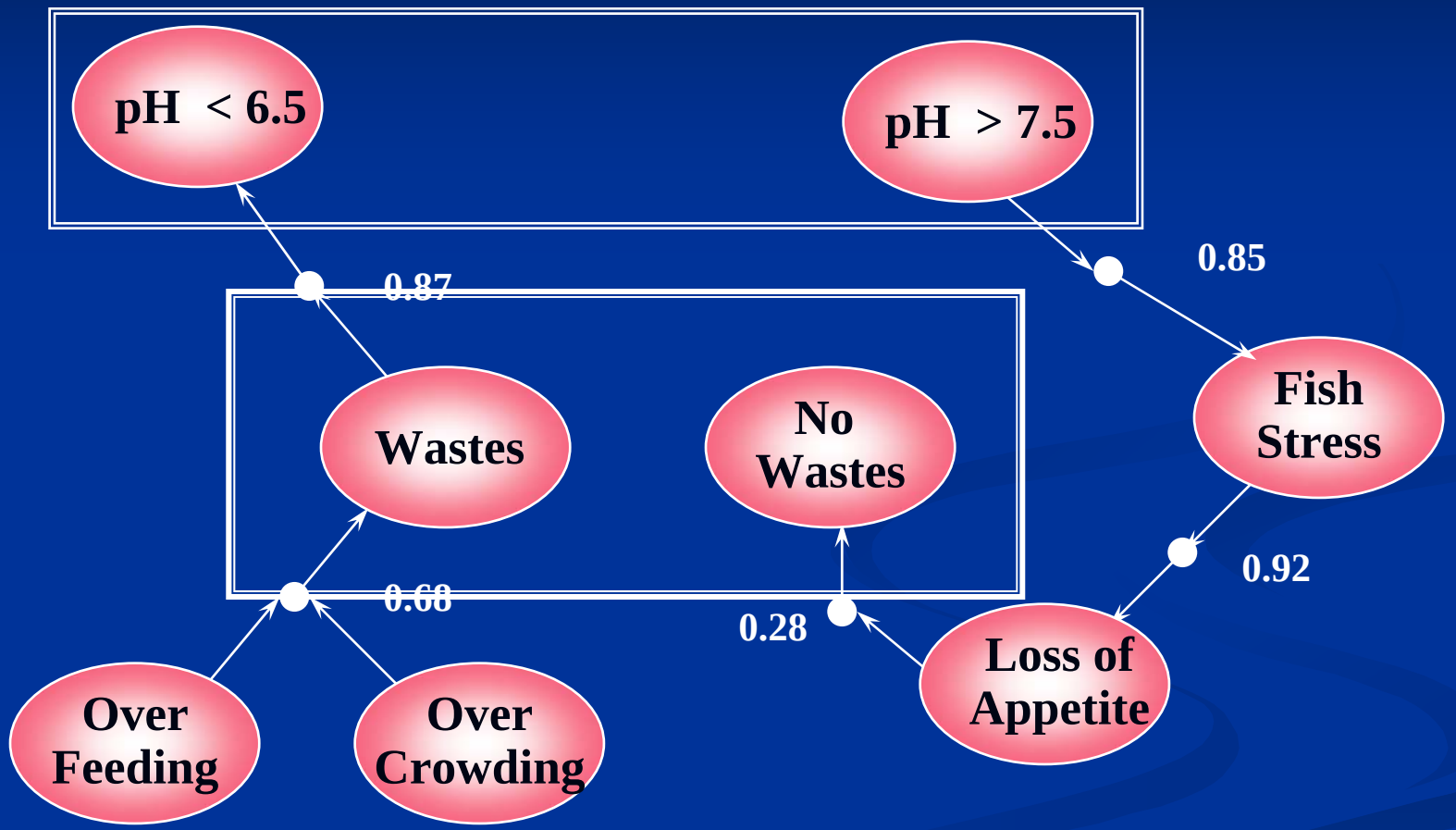
(Santos & Santos 96; Santos et al. 97; Santos et al. 97b; Santos & Santos 99; Shimony et al. 00; Johnson & Santos 00; Rosen et al. 01; Santos et al. 2004, Santos & Dinh 2008)

BKB Structure



- Directed graph
- Consists of instantiation-nodes (I-nodes) and support-nodes (S-nodes)
- I-nodes are r.v. states
- S-nodes capture probability info

Example BKB Fragment

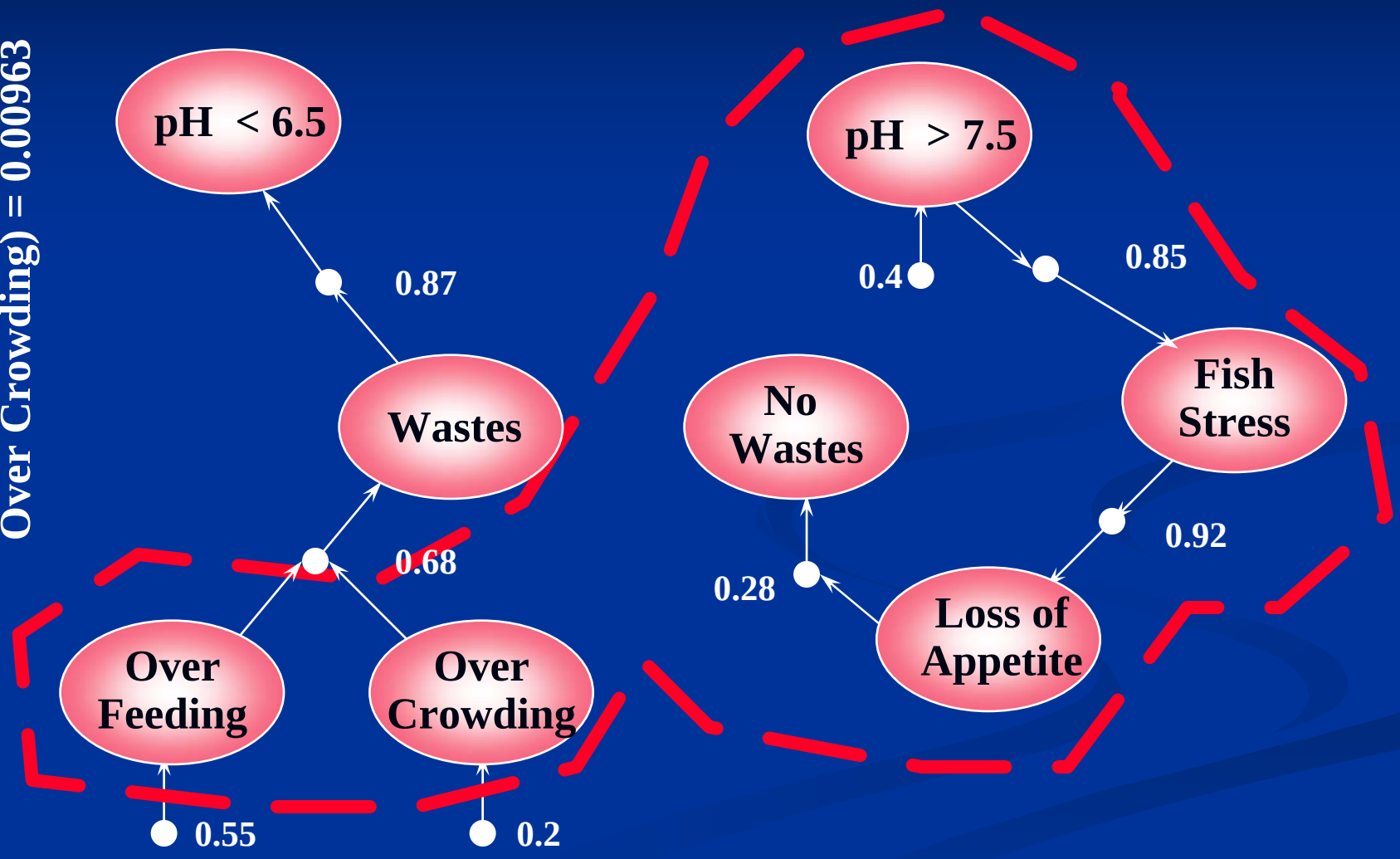


Joint Probability Example (Making Inferences)

- Treat I-nodes as logical XORs and S-nodes as logical ANDs
- Valid inference implies a valid truth assignment
- Inference is subgraph of BKB consisting of all the true nodes and their edges
- Each r.v. has at most one instantiation and is supported
  - Focus is on belief revision as opposed to belief updating
  - Searching for the most probable computable world/explanation that supports the evidence

Subgraph and Inference

P(pH>7.5, Fish Stress, No Wastes, Loss of Appetite, Over Feeding, Over Crowding) = 0.00963



Information Processing

- Types of queries on Bayesian Knowledge-Bases
  - Any queries of the form:
    - P(A1=a1, A2=a2, …, An=an | B1=b1, …, Bm=bm)

Probabilistic Consistency

- Given some rv A and rv sets α and β, if rules "α implies A" and "β implies A" are in the BKB, then <u>one</u> of the following holds:
  - α is inconsistent with β
  - α is consistent with β *and* probabilities of rules are equal
- Principle of *mutual exclusivity*

[Santos & Santos 96; Santos & Banks 96]

Probabilistic Consistency

- Given an set of rules:

  α_1 implies $A=a_1$

  α_2 implies $A=a_2$

  …

  α_k implies $A=a_k$

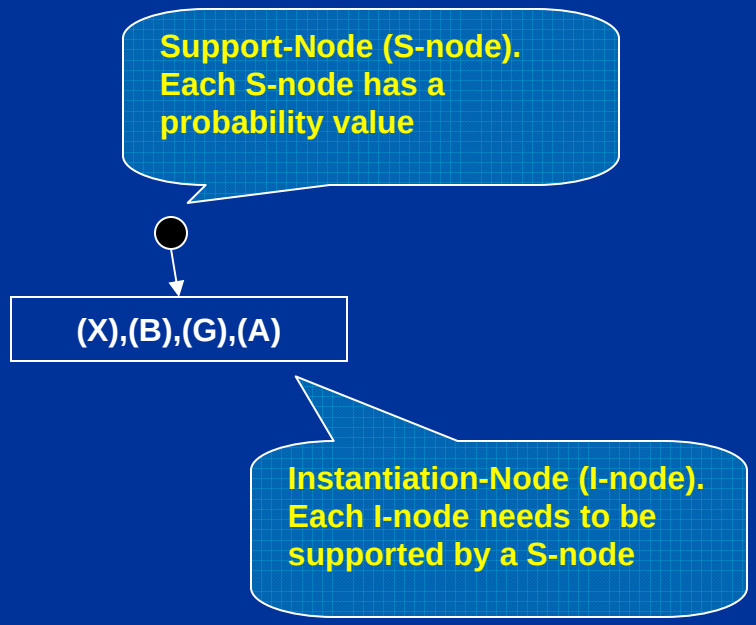
  where all the α_i's are mutually consistent and for all i, j, $a_i \neq a_j$ when $i \neq j$, the sum of the probabilities of this set is less than or equal to 1.0.

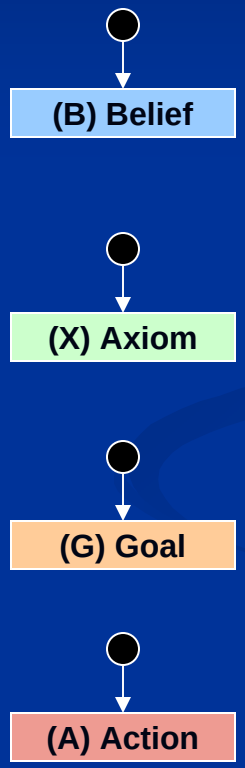
Probability Distribution

- Theorem: Given a BKB, the set of inferences derivable from the BKB defines a probability distribution over the variables in the BKB.
  - Each inference defines a unique joint probability
  - Incompleteness implies more than one possible
  - Unique default distribution derivable

[Santos & Santos 96; Santos & Santos 99; Rosen, Shimony, and Santos 01]

Basics for BKB fragments and Entity Intent Inferencing Model





What the entity believes about others

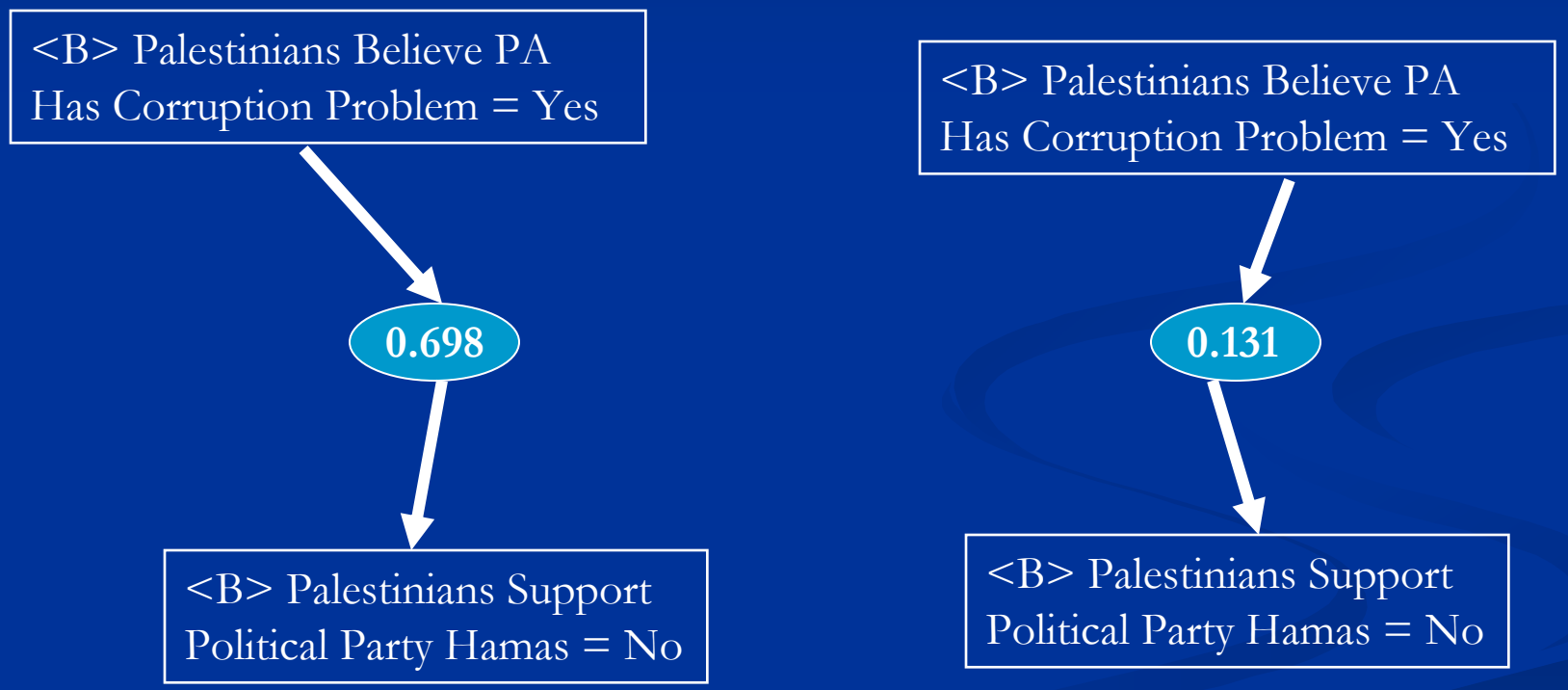
What the entity believes about themselves

What results the entity wants to achieve

How they will carry out their tasks

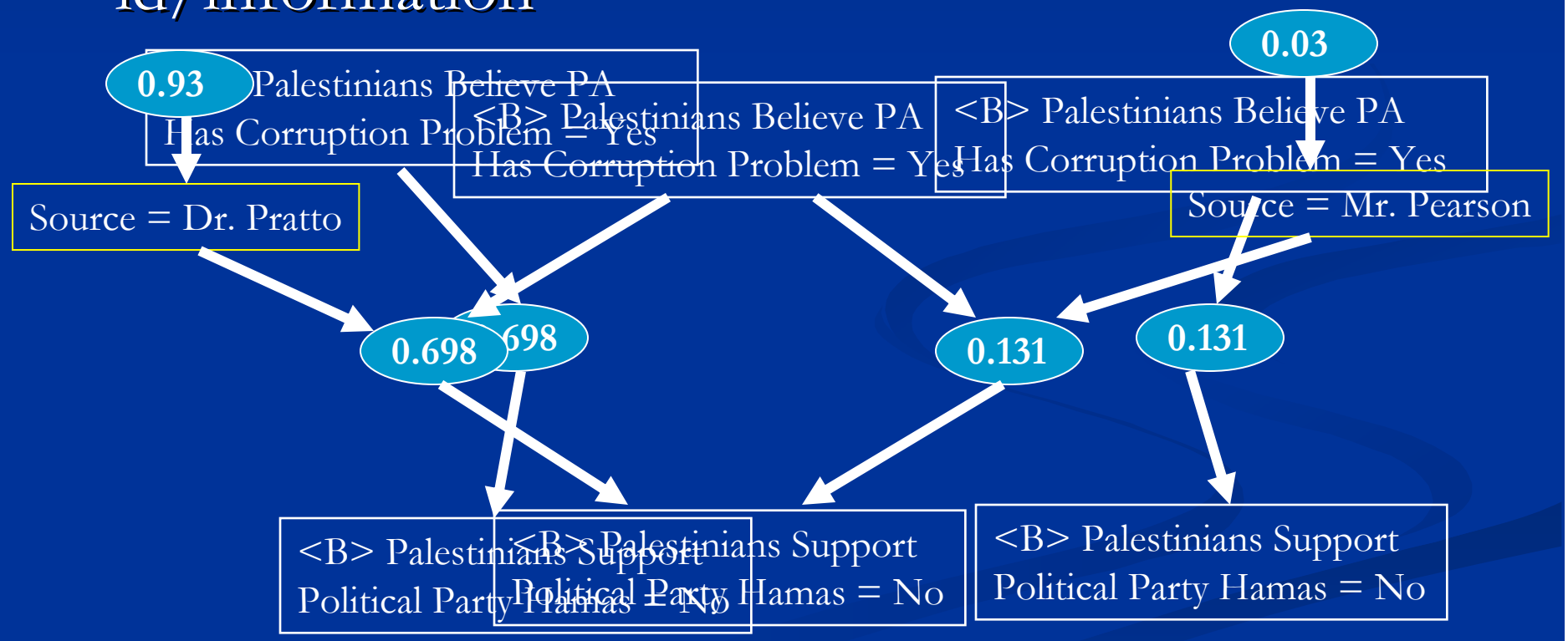
Aggregating BKFs

- Two experts have conflicting opinions/theories



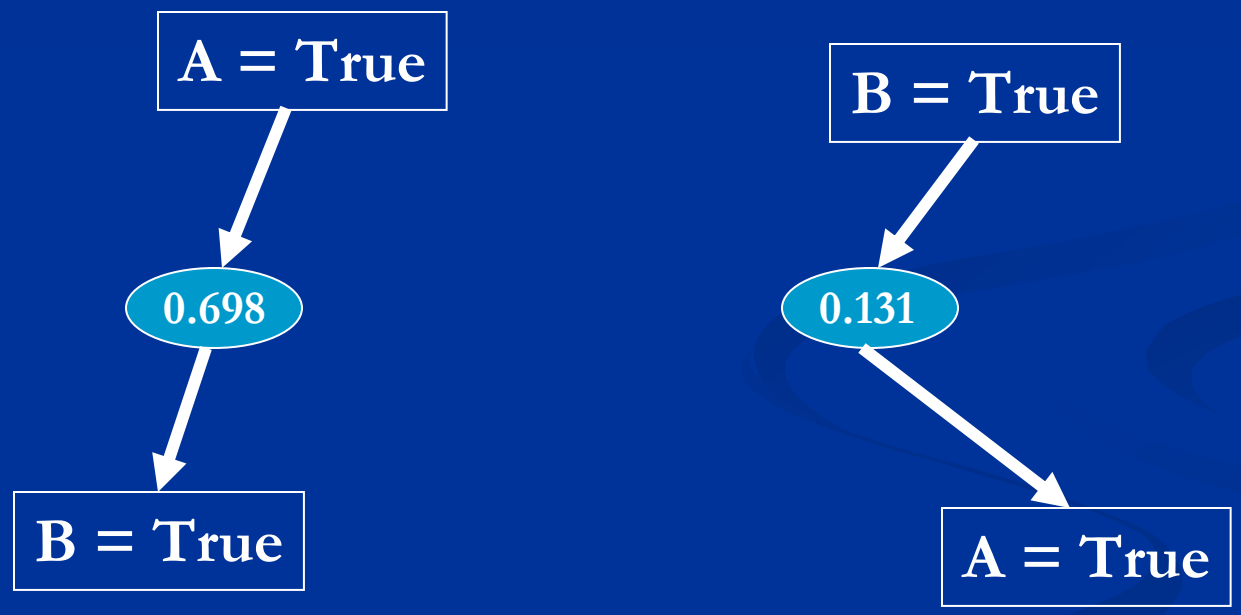
Source-Based Knowledge Fusion

- Basic Approach – tag fragment with source id/information

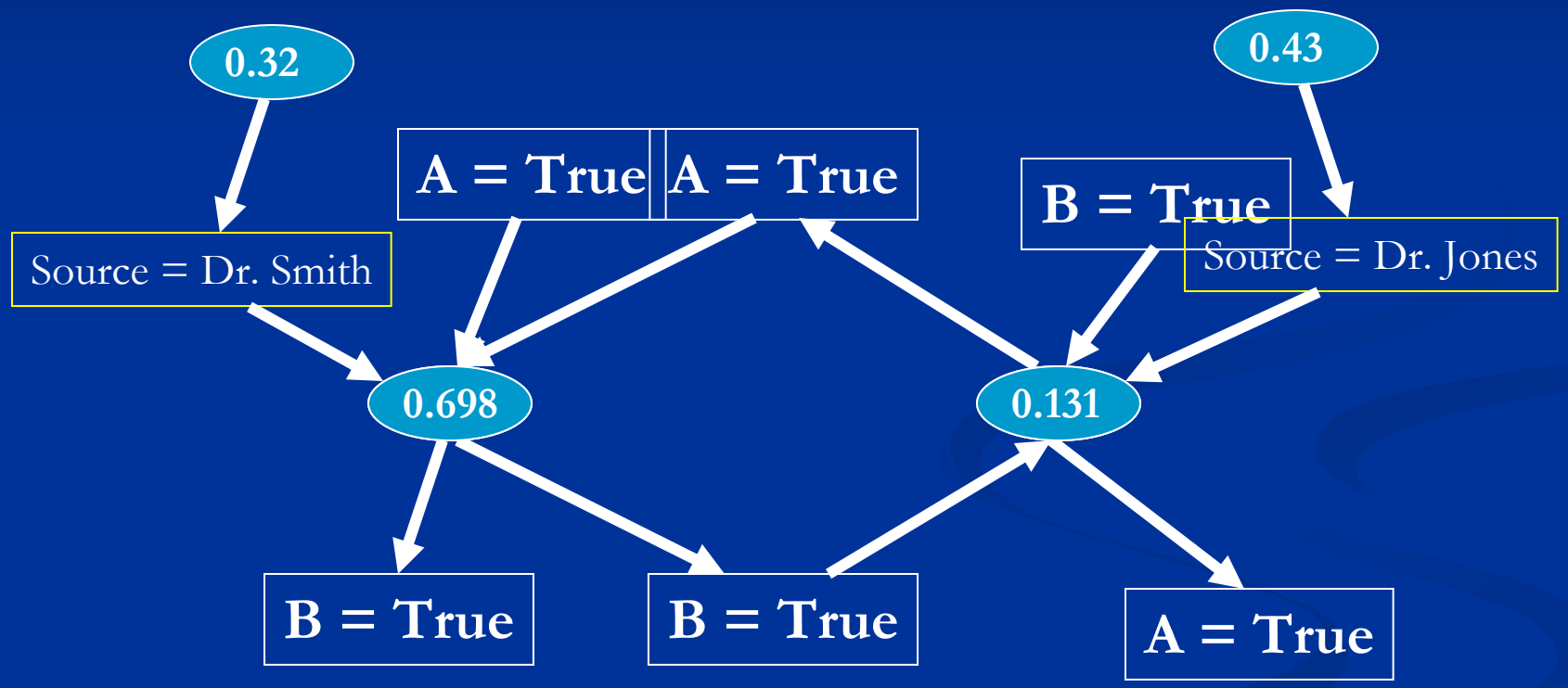


Loopy Problem

- Two experts disagree on causes and effects

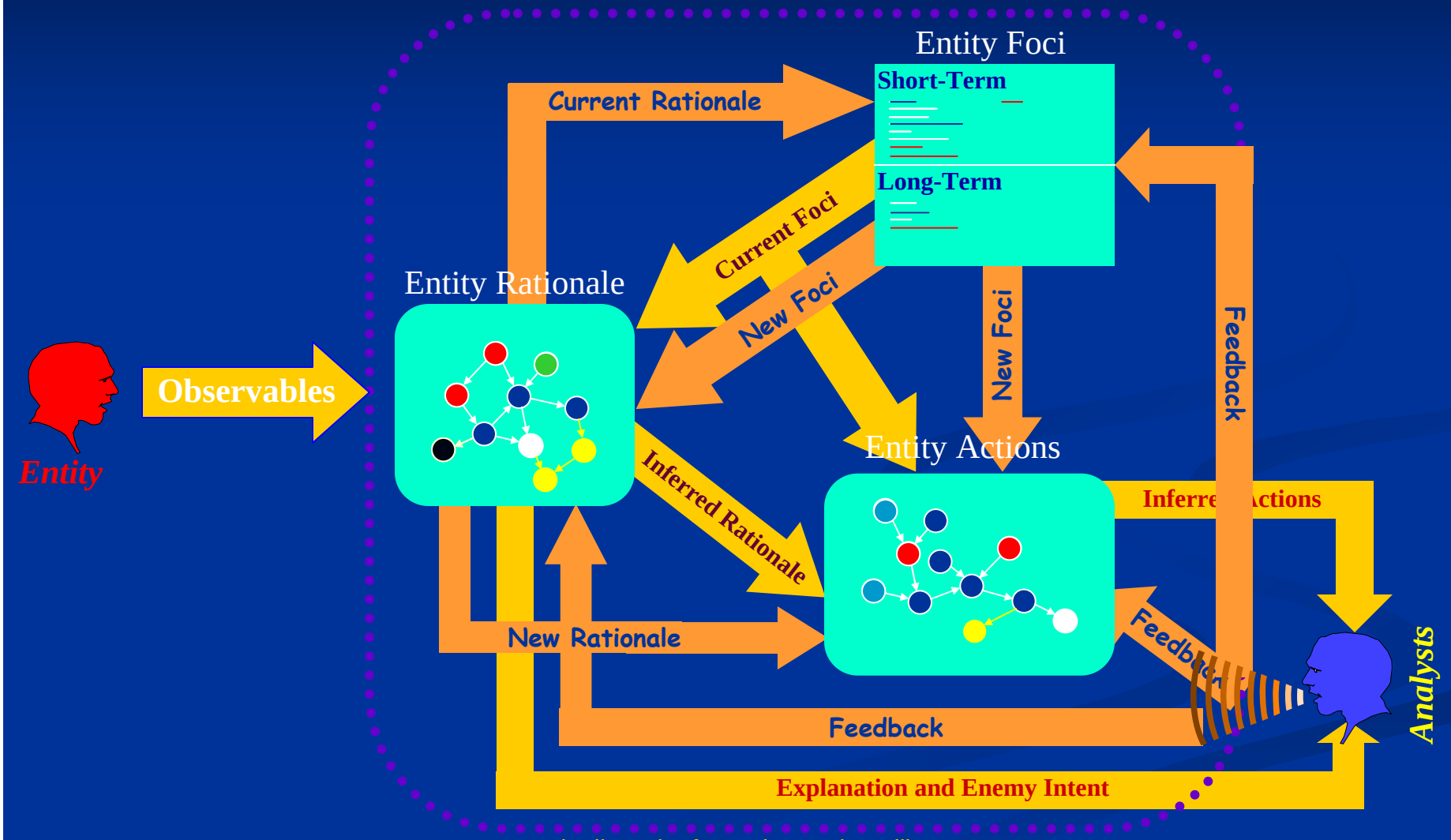


Loopiness?


0.32
0.43
A = True
A = True
B = True
Source = Dr. Smith
Source = Dr. Jones
0.698
0.131
B = True
B = True
A = True

Processing for Entity Intent

(Santos 2003; Santos & Zhao 2006)



Outline

- Traditional approach to control system resilience
  - Deficiencies of Traditional Approaches when Confronted by Intelligent Adversaries
- What is Intent?
- Modeling Intelligent Adversaries
- Examples of Intent Inferencing
- Conclusion and Discussion

Intelligence Support

- Getting the right information to the right place at the right time
- What other information available?

- Modelling Analysts Through Dynamic User Modelling for Novel Intelligence from Massive Data
  - Sponsored by ARDA/NIMD (completed)
  - Team members – Q. Zhao (Dartmouth), H. Nguyen (UWisc – Whitewater)
- Enabling a Collaborative Problem-Solving Framework Through User Intent Modeling of the Analytic Process
  - Sponsored by IARPA/CASE (ongoing)
  - Team member – H. Nguyen (UWisc – Whitewater)
- A Framework for Detecting Cyber Insider Threats
  - Sponsored by AFOSR (ongoing)
  - Team member - H. Nguyen (UWisc – Whitewater)

Insider Threat Detection

- Project Goal - Research and develop a framework for intent-driven insider threat detection [Santos et al '08]
- Focus – Users involved with information seeking tasks for intelligence gathering and analysis

Definitions

- Who are insiders?
  - Either previous or current members of an organization who have access to privileged resources
- Who are insiders in an intelligence workspace?
  - Either previous or current members of an organization whose intelligence analyses have impact on final decision making
- Who are malicious insiders in an intelligence workspace?
  - Insiders who have malicious intent to carry out attacks that decrease the performance of the group
    - E.g., impacts analytical output/quality and decision-making

Challenges

- Lack of effective measurements
  - Damage done by insider attacks are covert
    - E.g. Hide/Alter critical information
  - Observables are difficult to measure
    - E.g. Quality of reports/analyses
- In analytical systems, the detection methods are required to be flexible
  - The activities analysts can carry out to achieve their goals are not constrained
- Intention of the damage is hard to detect
  - Misinformation vs Disinformation

Approach

- Develop a computational model for detecting abnormal behaviors of an analyst via capturing and analyzing an analyst's intent in an analytic process and inferring his/her behaviors.
- The model is constructed based on observations of internally measurable factors such as his/her hypotheses, information gathered/read/written or printed by an analyst and evidence gathered from intelligence sources.

Deception

- Current research does not take into account the impact of *deception* in insider attacks
  - Use deception to access/alter critical data
  - Use deception to carry out malicious actions
  - Use deception to divert attentions
  - Use deception to hide critical evidence
- Need deeper understanding about how deception is involved in the incident
  - helps alert insider threats
  - helps detect malicious insiders after attacks occur
- Little available research focused on insider threats in information seeking/access domain for intelligence gathering and analyses

What Is Deception?

- Definition
  - *Information designed to manipulate the behavior of others by inducing them to accept a false or distorted presentation of their environment- physical, social, or political (Whaley, 1991)*
  - *Deliberate act perpetrated by a sender to engender in a receiver's beliefs contrary to what the sender believes is true to put the receiver at a disadvantage (Burgoon, 1994)*
- Classification (Whaley, 1991)
  - Simulative deception: create false beliefs
  - Dissimulative deception: hide the truth

Deception Detection using Simulation Tools

- Simulation tools: software used to assist people in decision making
- Multi-agent system
  - program that models a group of experts
  - some contributing experts may not operate in good faith
- Adversary intent inference model
  - network that models an adversary in war-gaming environment
  - military tactics involve deception operations

- Motivation:

  Deception detection is challenging but important

Deception Detection in Multi-agent System

- Idea:

  Assume that all agents share similar knowledge, then each agent's predicted opinion is similar to its actual opinion.
- Methodology:
  1. Compare agents based on past opinion
  2. Simulate deception of one agent
  3. Predict opinions based on the correlations and other agents' opinions

  If an agent's predicted opinion is highly different from its actual opinion, then a possible deception is detected.

Experimental Evaluations

- ANOVA Analysis

Parameter	*Agents*	*Perturbation*	*Evidence*	*# times STD*
F Value	2.5	346.3	1524.04	1845.03
F Critical	3.1	2.68	3.1	2.68
P Value	0.09	<0.0001	<0.0001	<0.0001
Significance	Not	Yes	Yes	Yes

- Parameters that influence the detection rate
  - Perturbation value:
    the more knowledge shared, the more similar opinions
  - Number of pieces of evidence:
    the more information we have, the easier to detect unusual observables
  - Number of standard deviation:
    the fewer error we accept, the higher chance to identify deception
- Parameter that does NOT influence the detection rate
  - Number of agents:

Assumptions About the Decision Making Process

- An analyst should start out with many ideas or hypotheses, and as they gather more information their focus should narrow. The more information they have the more ideas they can dismiss.
- Analysts, especially ones with experience, will have bias. This bias should not be confused for an insider threat.
- The less a person knows about a subject, the more likely they are to change their mind. Less experienced analysts are more likely to flip flop between ideas (i.e. APEX_L).
- More information does not mean better results. If an analyst has fewer sources than the median of all the other analysts but they look at it in-depth compared to the other analysts, they may be making better use of their information.
- The goal of an analyst should be to disprove their current hypothesis, not to only look at information that supports it. Not reporting any information that goes against their hypothesis may be an indicator of an insider.

Assumptions About the Decision Making Process Cont.

- In critical thinking, "quantity leads to quality" (Heuer 77). Focusing on fewer ideas means that an analyst may miss an alternative they could have seen by thinking outside of the box.

- It is important to have a discernable difference between an unproved hypothesis and a disproved hypothesis. Having no information that proves a hypothesis is different than having information against a hypothesis.

APEX Data Set

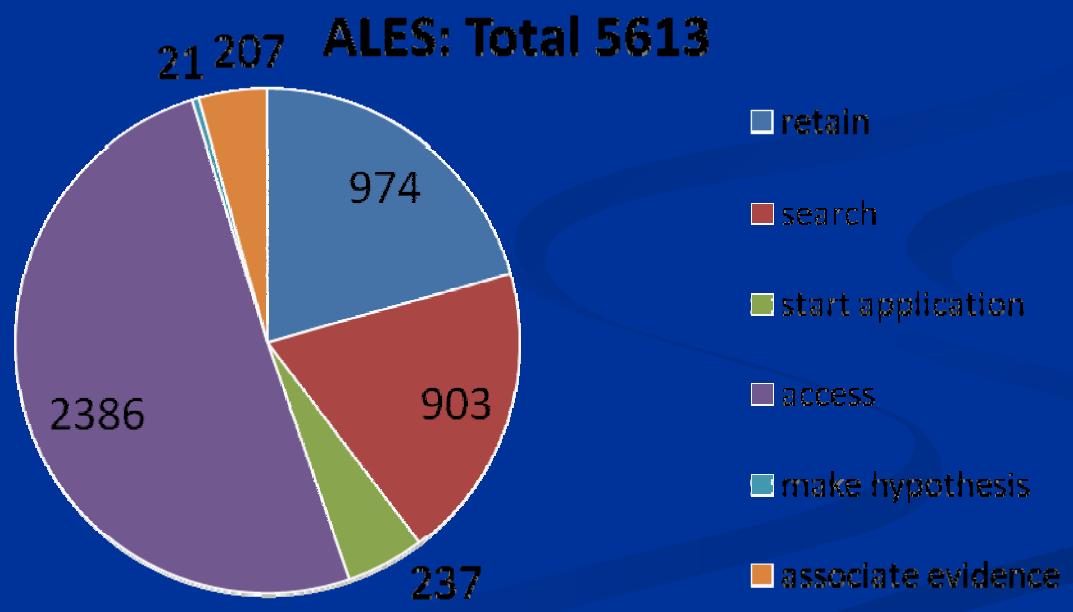
APEX 07 data set:
Data collected by *NIST* to evaluate CASE tools for tactic collaboration.
Scenario of the experiment: The Secretary of State has requested analytic assessment of these two questions, in priority order:

1.Where does the Iranian clerical community stand on Ayatollah Khamenei and President Ahmadinejad's policies with regards to Iran's civilian and military nuclear program?
2. Are there fissures in the clerical community and do they represent a deepening divide among the clerics loyal to the Iranian revolution?

APEX Data Set

APEX 07 data set:

- 9 users (including root for testing purpose),
  - APEXB, APEXC, APEXE, APEXF, APEXH, APEXL, APEXK, APEXP, and root



Taxonomy of malicious actions in intelligence analysis

Search Queries

- Put more constraints when searching nonsupport queries
- Search more support queries than nonsupport queries

Save Information

- Use outdated documents when support documents are not sufficient
- Ignore nonsupport documents
- Create Fake Snippets
- Unbalanced quantity of support and nonsupport documents

Draft Reports

- Over cite same documents
- Exaggerate/Alter Evidence Ratings
- Use Fake Snippets as evidences
- Quote content out of context

Design of Malicious insiders

- Three analysts among 11 are malicious insiders
  - Musadegh
    - A senior information analyst
  - May
    - A senior DIA analyst
  - Joe Nilson
    - A junior intelligence analyst
- Insiders are each derived from one of eight original analysts in APEX 07

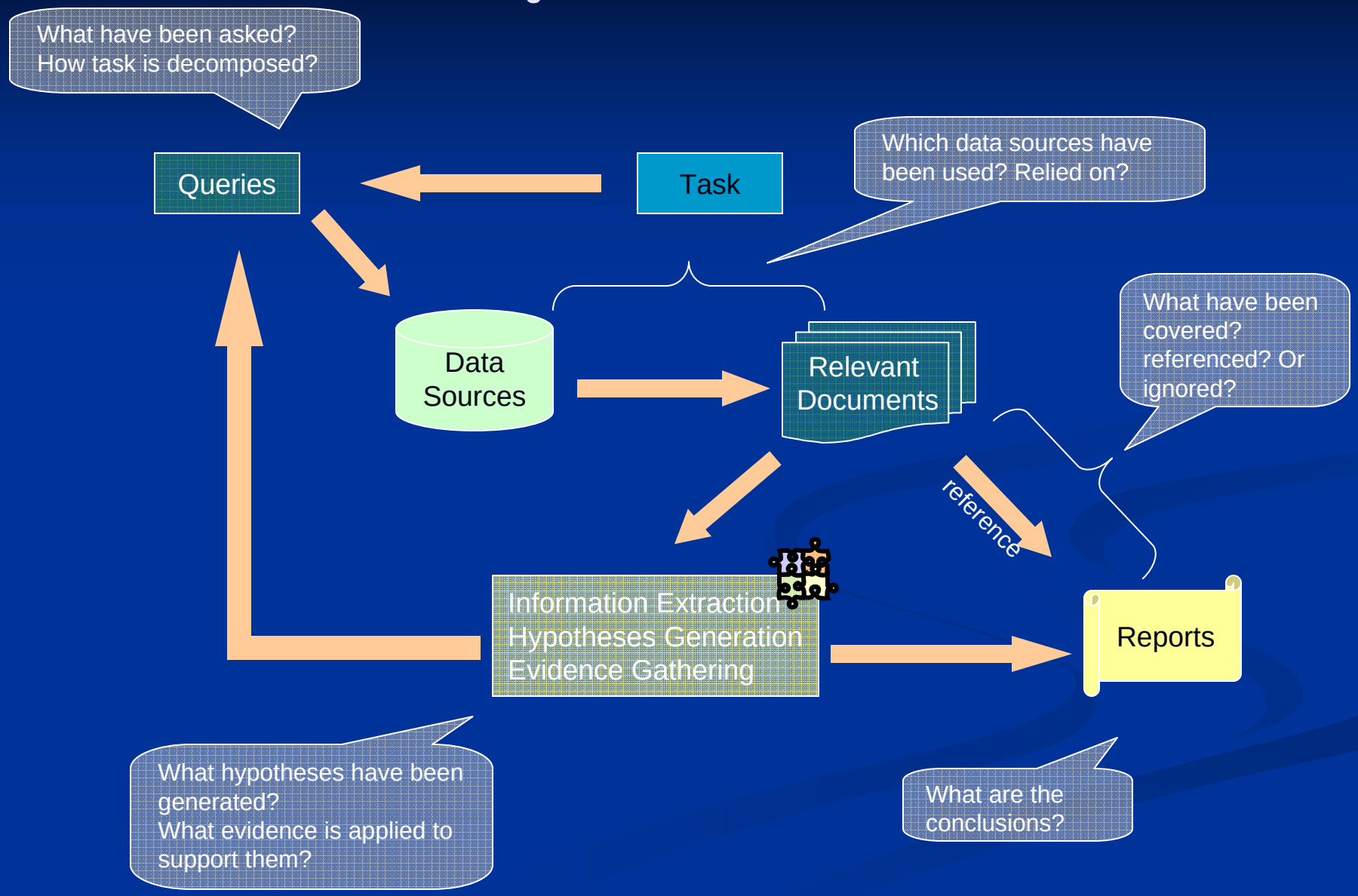
An Intent Modeling Approach

- Observe the analyst's activities and analyze the information they are accessing
- Intent driven
- Model the decision making process based on how "other's" views the world
- Build network fragments for each piece of information / knowledge, and merge them together for reasoning
- Based on Bayesian Knowledge Bases (BKBs) [Santos & Santos 97]

Analyst Intent Model

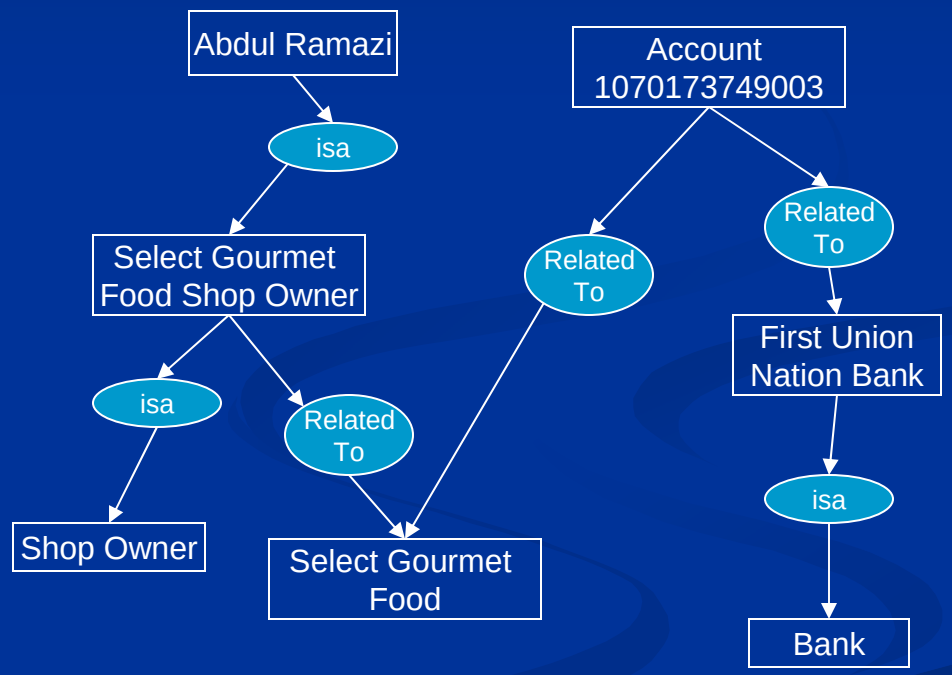
- ***Analyst intent*** = Goals + Actions + Commitment
- Our user model consists of 3 components that are designed to capture intent:
  - ***Foci***: *"**What** is the working space of the analyst and what they are concentrating on"*
  - ***Rationale***: *"**Why** does the analyst have these foci?"*
  - ***Action***: *"**How** are the analyst's goals accomplished?"*

Analytic Process



Document/Query Graph (an example)

Abdul Ramazi is the owner of the Select Gourmet Foods shop in Springfield Mall. Springfield, VA. (Phone number 703-659.2317). First Union National Bank lists Select Gourmet Foods as holding account number 1070173749003. Six checks totaling $35.000 have been deposited in this account in the past four months and are recorded as having been drawn on accounts at the Pyramid Bank of Cairo, Egypt and the Central Bank of Dubai, United Arab Emirates. Both of these banks have just been listed as possible conduits in money laundering schemes.

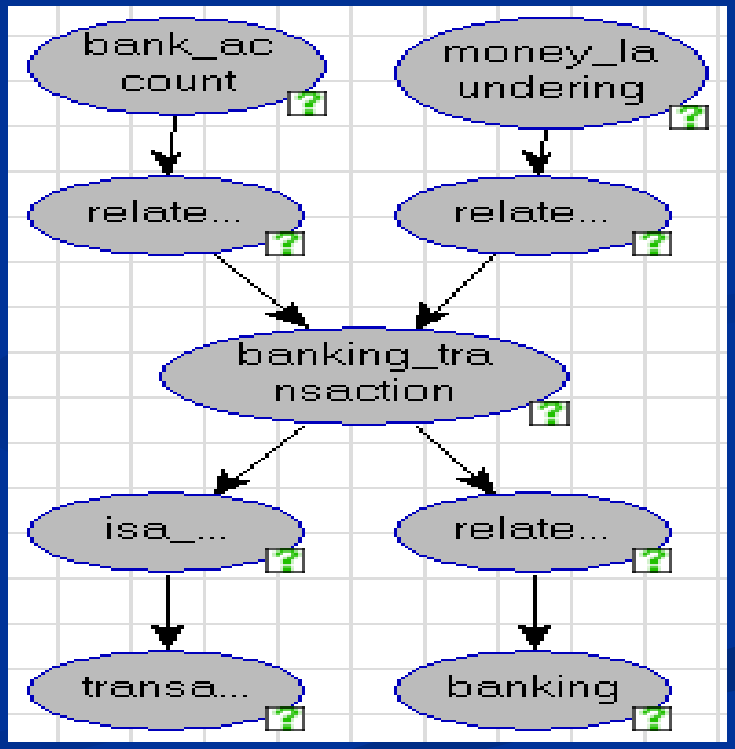


Foci (Interests)

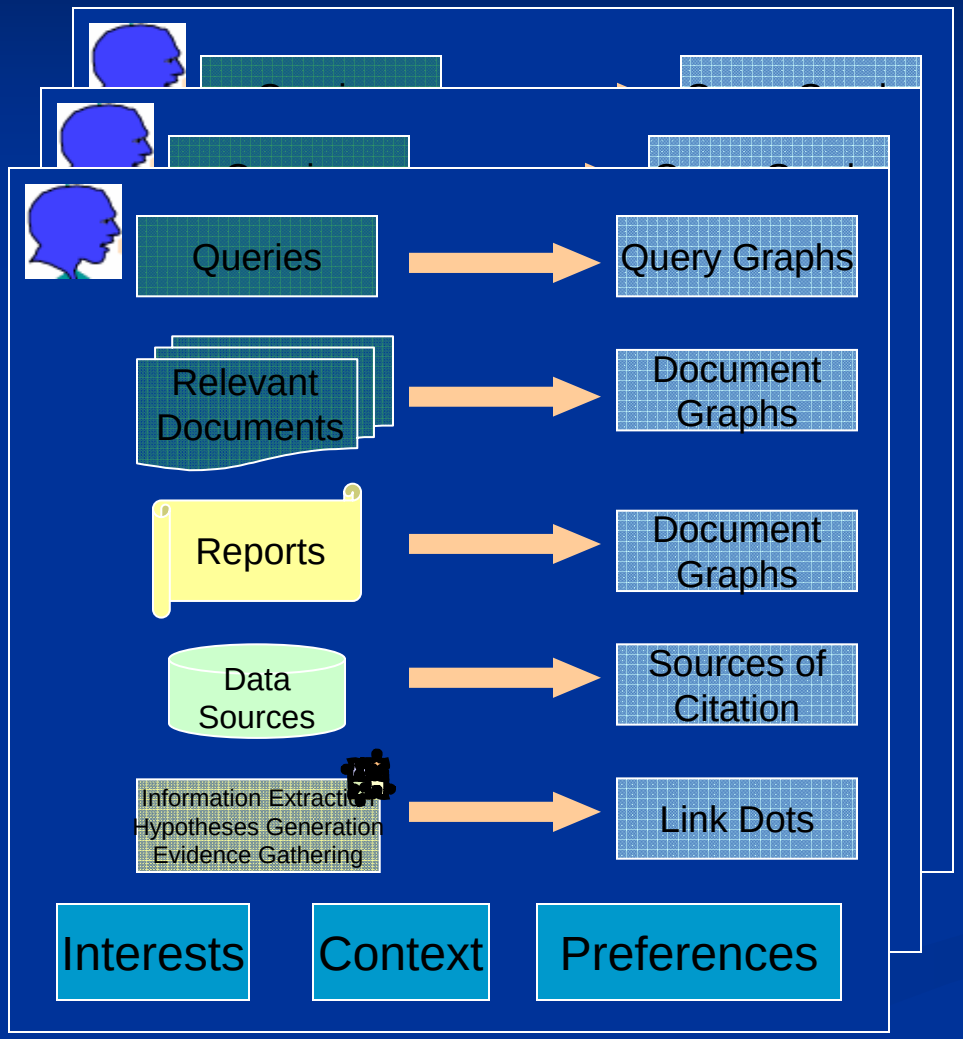
- What: Vector representing topics and goals of interests. Each element contains two attributes: topic of interests, and weight for that topic
- How to compute Foci: union of the following sources:
  - The analyst' current query
  - The analyst's queries in the past
  - Related documents: relevant documents in information seeking, document the analyst read, print, write, and reference.

Create and update rationale (context) network

- What: relations among topics of interests (high level and low level).
- Network is constructed "on-the-fly" by finding a common set of sub-graphs among the relevant documents
  - Captured as BKB
- Each document is represented as a document graph (DG).
- A sub-graph X of DG Y is:
  - X is a DG
  - $\forall$ node $a \in X \rightarrow a \in Y$



Overall architecture



- The same task has been assigned to several analysts. Each analyst has a user model representing his/her
  - Queries
  - Related documents (documents read/written/cited/printed /referenced by the analyst)
  - Hypotheses
  - Context
  - Interests
  - Preferences

Hypotheses

- Malicious intent can be inferred based on a series of actions taken by an insider
- Group of insiders working on similar tasks can form an external reference set for inferring malicious intent
  - Used to detect inconsistency between different analytic stages
- Serial user models for each insider in time order also forms an internal reference set
  - Used to detect change of mind/topic/interests

Detection based on Clustering

- Hypothesis: a malicious insider behaves inconsistently
  - Between query and report
  - With normal insider on information assessment
- Methodology:
  - Cluster a group of analysts based on query and report, compute consistency value between query cluster and report cluster
  - For each analyst, look for the analysts who cite the same evidences, calculate the correlation between their corresponding ratings

Consistency with other Analysts on Information Assessment

- Step 1: look for the same evidences from each pair of analysts' reports belonging to the same group in the assessment cluster

Hypothesis	Url	Ratings from APEX_PParam	Ratings from APEXK
1	http://129.6.84.47:18080/cmsRest/cms_rest?id=495cce1bX11635035eb9XY1a68	-1	0.53333
1	http://129.6.84.47:18080/cmsRest/cms_rest?id=495cce1bX11635035eb9XY29e8	-1	0.8
1	http://129.6.84.47:18080/cmsRest/cms_rest?id=495cce1bX11635035eb9XY6a82	-0.25	0.5
2	http://129.6.84.47:18080/cmsRest/cms_rest?id=495cce1bX11635035eb9XY1a68	1.333333	0.45
2	http://129.6.84.47:18080/cmsRest/cms_rest?id=495cce1bX11635035eb9XY6e50	2	1.5
2	http://129.6.84.47:18080/cmsRest/cms_rest?id=495cce1bX11635035eb9XY29e8	2	0.375
2	http://129.6.84.47:18080/cmsRest/cms_rest?id=495cce1bX11635035eb9XY6a82	0.333333	-0.25

- Step 2: Calculate correlation between each pair of analysts' assessment on the evidence

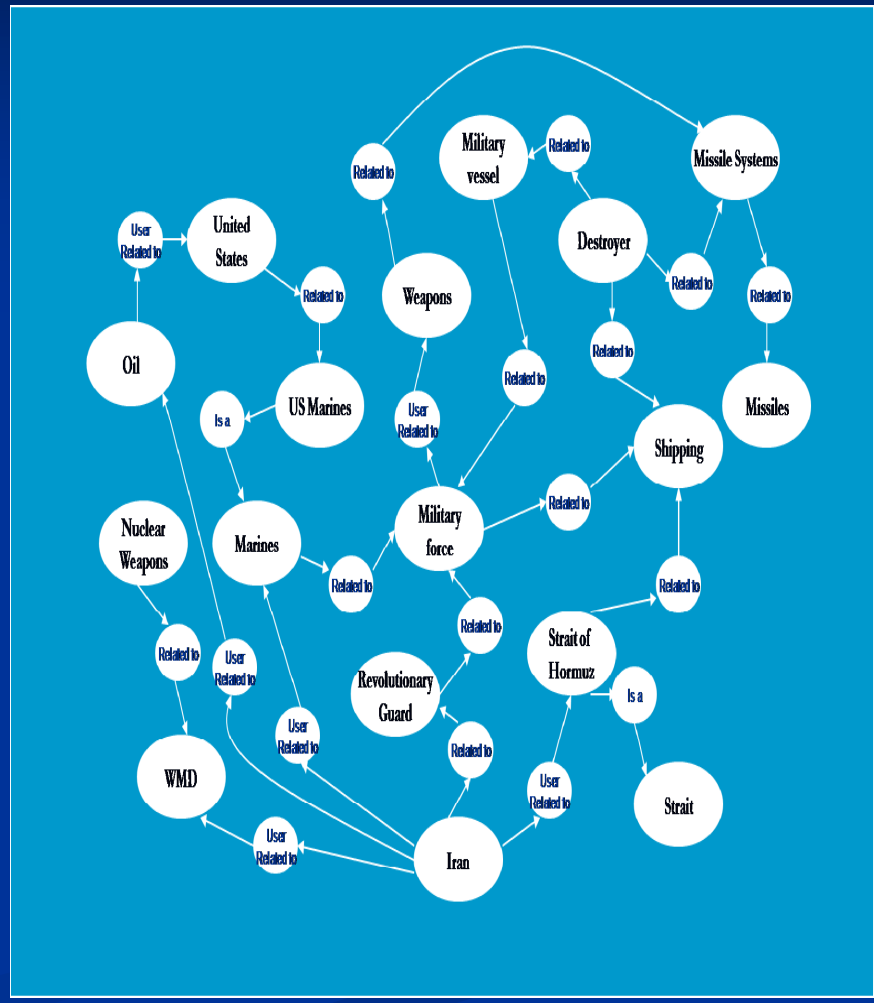
  e.g. Correlation between APEXPParam and APEXK = 0.219

Current Clustering Results and Conclusions

- Using clustering methodology alone doesn't capture malicious insiders effectively
  - Clustering is based on keywords so it doesn't cluster the insiders according to the semantic meaning of the data
  - Clustering tends to be insensitive to change with a small number of insiders
- Methodology based on clustering likely to capture novice insiders

Detection based on User Modeling

- User Modeling is one of the techniques that captures insiders' behaviors, interests, and knowledge bases
- User Modeling can be built given the data an insider has read, written or executed in a time order
- User Modeling can be used to detect changes of topics/interests over time, to recognize whether a document is of interests, and even predict an insider's goals and intentions.
- Therefore, user models allow us to measure and analyze different insiders in order to understand, anticipate, and detect insider attacks.



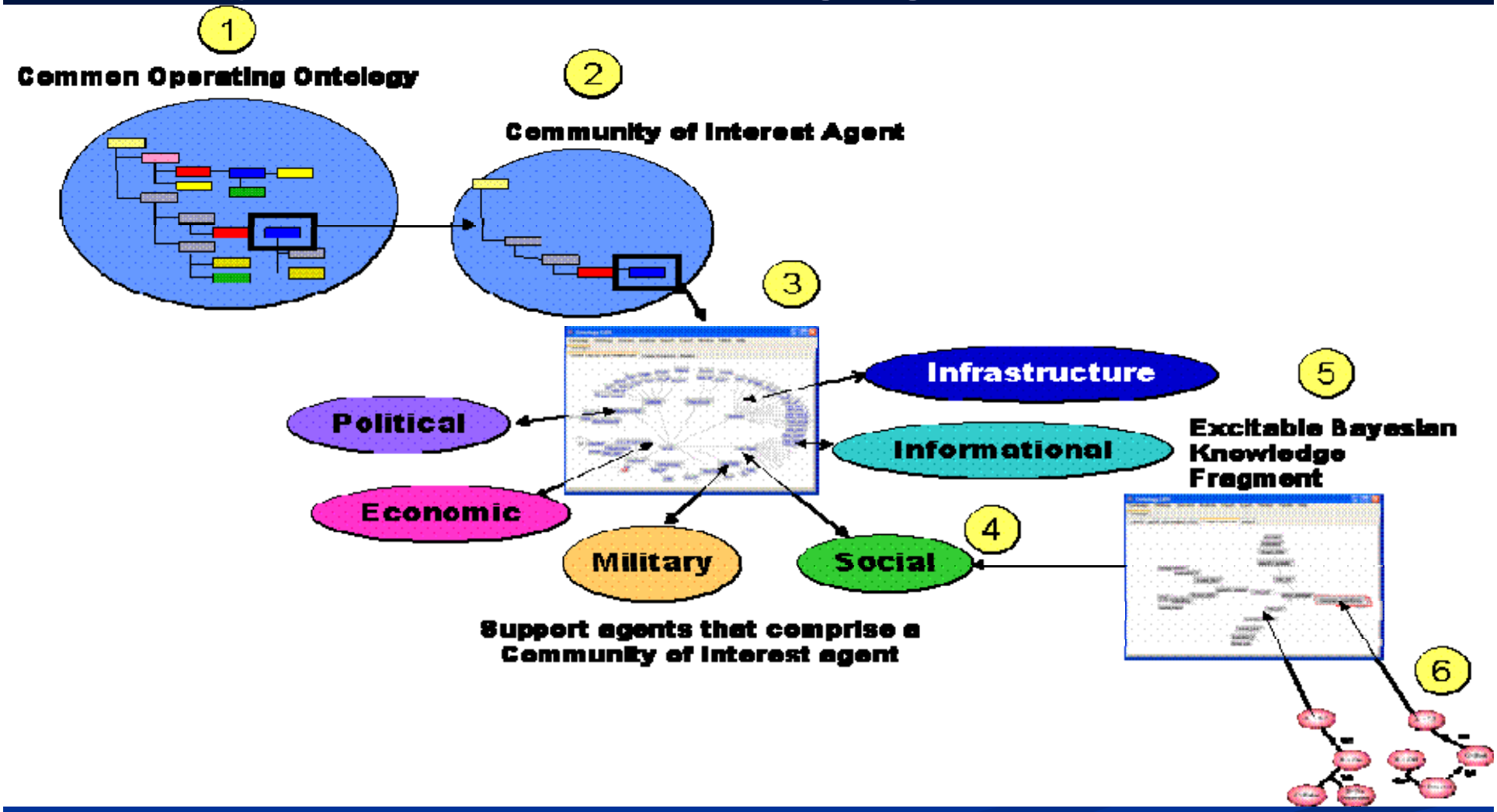
Insider Detection Using Intent Modeling

- Experimental setup
  - Build user models given a set of recorded actions for each analyst (+ insider)
  - Different experiments on user models
    - Serial autocorrelation between consecutive user models in time order
    - A series of similarities between user model at each step and final reports
    - Cross comparisons for a pair of user models
    - Compare changes of user models between different types of actions

Discussion

- Intent modeling for insider threat detection with current approaches provided more insights into problem
  - Testbed insiders are "expert" quality
    - Currently hard to distinguish from original analysts
  - Selection of evidence to support assessment needs more analyses (better understood)
  - With good analysts (even insiders), user models will converge with final assessment reports
  - Need to explore more similarity metrics beyond basic graph isomorphism
  - (See forthcoming IAT 2008 paper for new results).

Dynamic Adversarial Gaming Algorithm (DAGA)



Team: Securboration, Inc. and Dartmouth College

Sponsor: AFOSR Phase II STTR [FY 07-08]

Program Overview

- *Customer: AF Office of Scientific Research*
- *Contract Duration: Dec 2008*

- *AFOSR Focus Area*
  - *Develop algorithmic techniques to accurately predict Community of Interest (COI) responses to social, cultural, political and economic actions.*
    - *Enable predictions based not only on current situation and adversary capabilities, but also on adversary's cultural dimensions and 'soft-factors'.*
    - *Use predictions to provide adaptive strategy selection in multi-cultural adversarial games and related simulations within the context of an agent-based dynamic adversarial environment.*

Operational Need

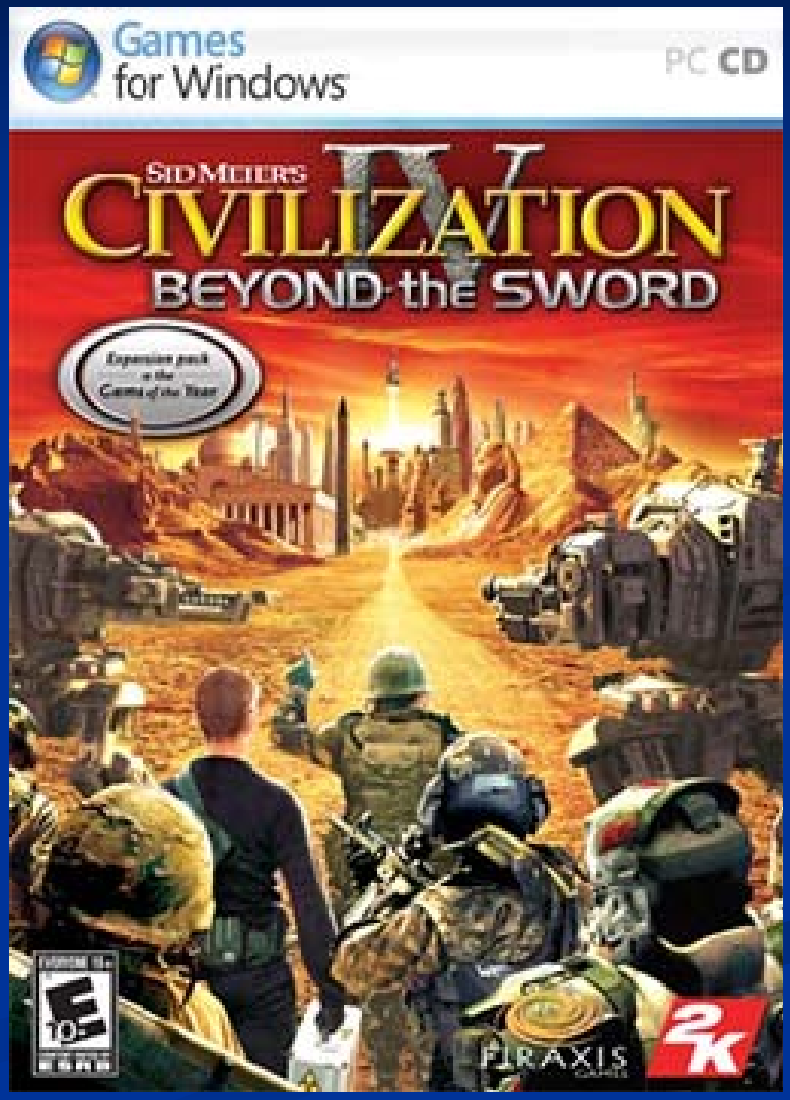
- Realistic, dynamic adversaries modeling capability
  - Asymmetric, adaptive adversary for wargaming and mission rehearsal
  - Added realism for training, planning, and threat detection
- Provide real world adversarial behavior for simulations
  - Supports the move away from doctrine based warfare on the part of an adversary towards more realistic asymmetric response
- Show both internal and external influences affecting adversary behavior

Operational Applications

- Initial focus on Gaming with transition to areas such as
  - Asymmetric Threat Detection
  - Mission Planning
  - Counter-terrorism
- Fundamental capability of DAGA is to predict individuals or group response to social, cultural, political and economic actions
  - Homeland Security / Intelligence
  - Potential acts of terrorist cells

Game Integration

- To highlight DAGA's capabilities, we have integrated it with the popular *Civilization 4* (Civ4) game engine to demonstrate how the infusion of socio-cultural influences leads to a much more realistic asymmetric adversary.

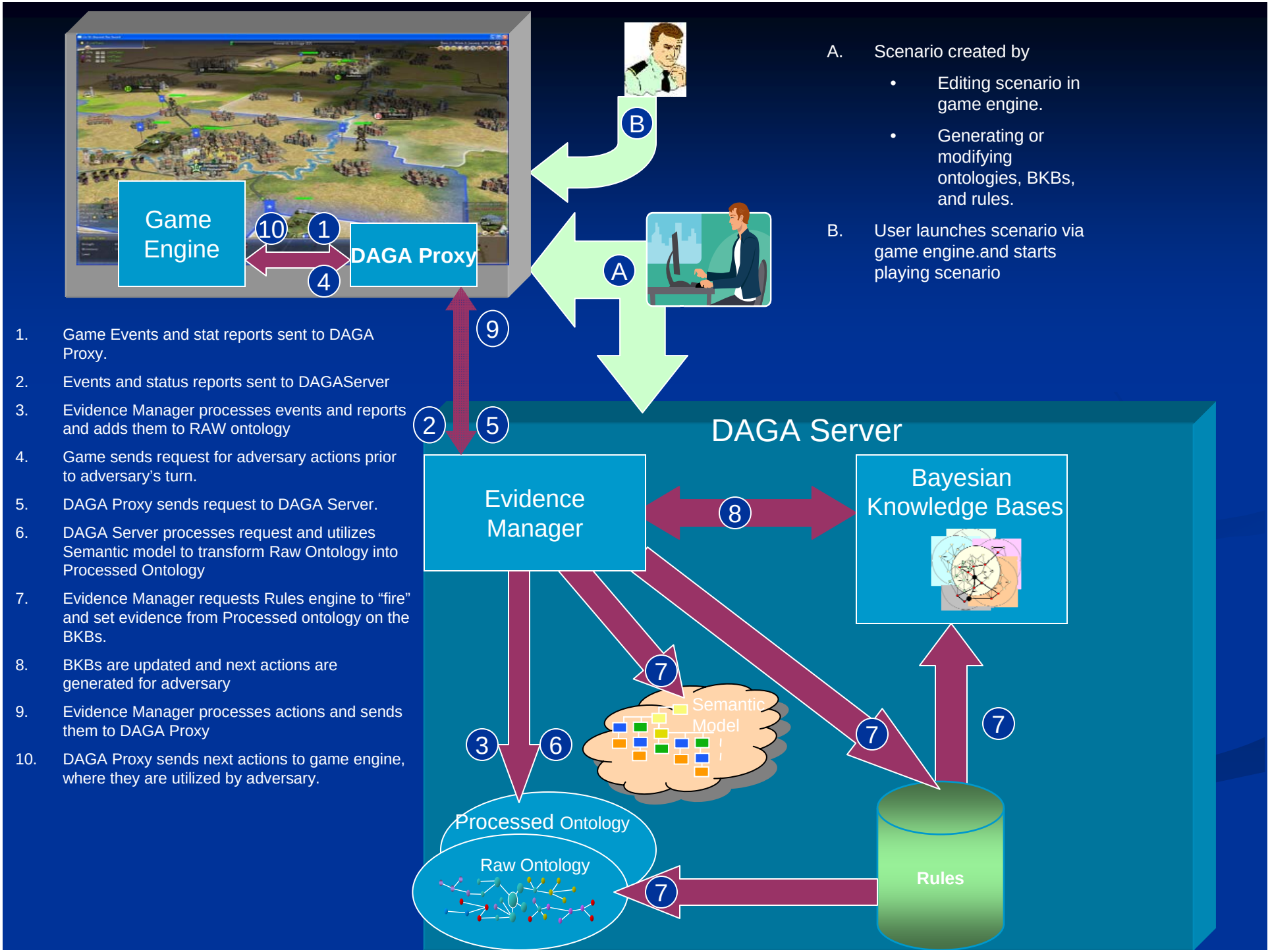


Game Scenario

- Developed scenario representative of the current political and military situation in Baghdad
  - "Players" include **Coalition Forces**, Iraqi Transitional Government, Mahdi Army, Al Qaeda in Iraq, and Ansar Al-Islam.
  - Each player is represented as a Community, with their own goals, actions, beliefs, and axioms which are modeled as Bayesian Knowledge Bases.
  - As the 'game' progresses, DAGA 'pulls' information from the gaming engine for use in its calculations, and 'pushes' results back to the gaming engine to dynamically modify the behavior each adversarial player.

Game Results

- The result is a game that now includes realistic asymmetric adversaries that act, and react to coalition actions, based on socio-cultural beliefs and other soft-factors.
  - Without DAGA, adversaries give up quickly because of overwhelming coalition force. With DAGA adversaries are more dynamic and continue to fight.



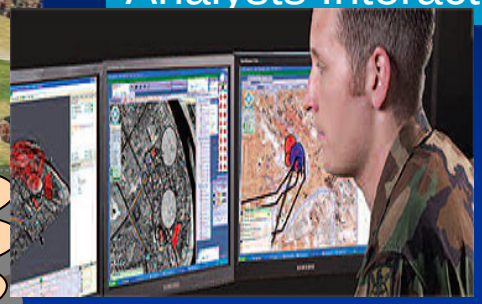
A. Scenario created by
- Editing scenario in game engine.
- Generating or modifying ontologies, BKBs, and rules.

B. User launches scenario via game engine.and starts playing scenario

1. Game Events and stat reports sent to DAGA Proxy.
2. Events and status reports sent to DAGAServer
3. Evidence Manager processes events and reports and adds them to RAW ontology
4. Game sends request for adversary actions prior to adversary's turn.
5. DAGA Proxy sends request to DAGA Server.
6. DAGA Server processes request and utilizes Semantic model to transform Raw Ontology into Processed Ontology
7. Evidence Manager requests Rules engine to "fire" and set evidence from Processed ontology on the BKBs.
8. BKBs are updated and next actions are generated for adversary
9. Evidence Manager processes actions and sends them to DAGA Proxy
10. DAGA Proxy sends next actions to game engine, where they are utilized by adversary.

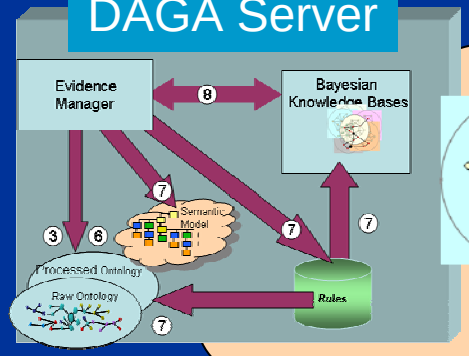
Simulation / Gaming Environment

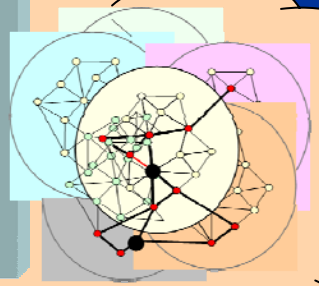
Analysts Interaction



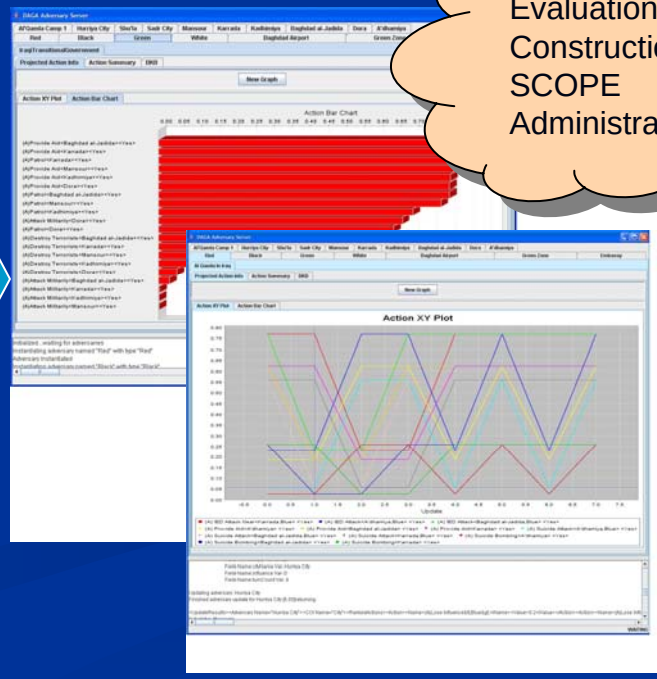
Real-time assessment, shifts in underlying cultural values based on actions and influences, feeding real-time operational planning systems.

DAGA Server



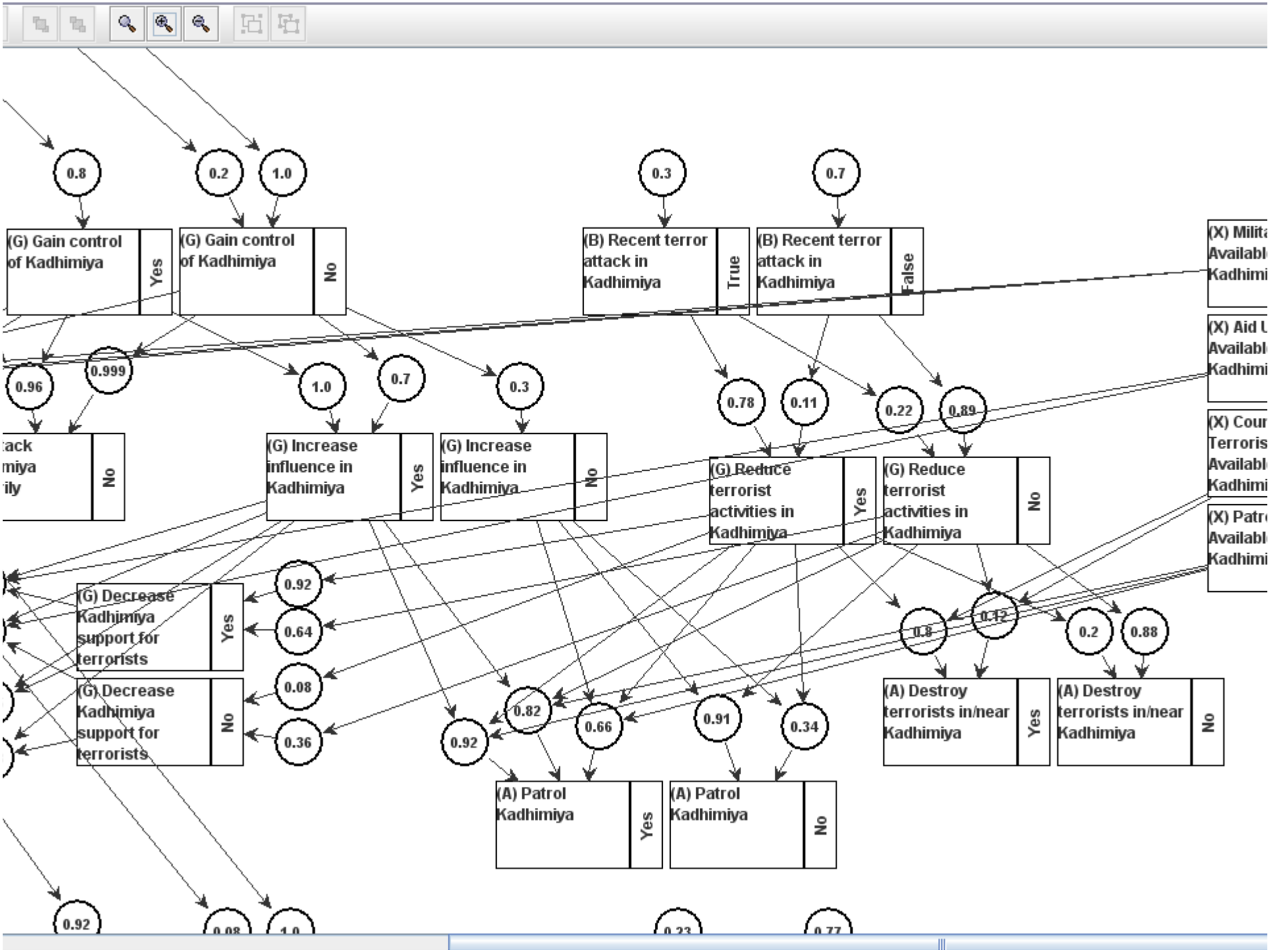


DAGA Computational Model



Model Validation, Evaluation, Construction, SCOPE Administration

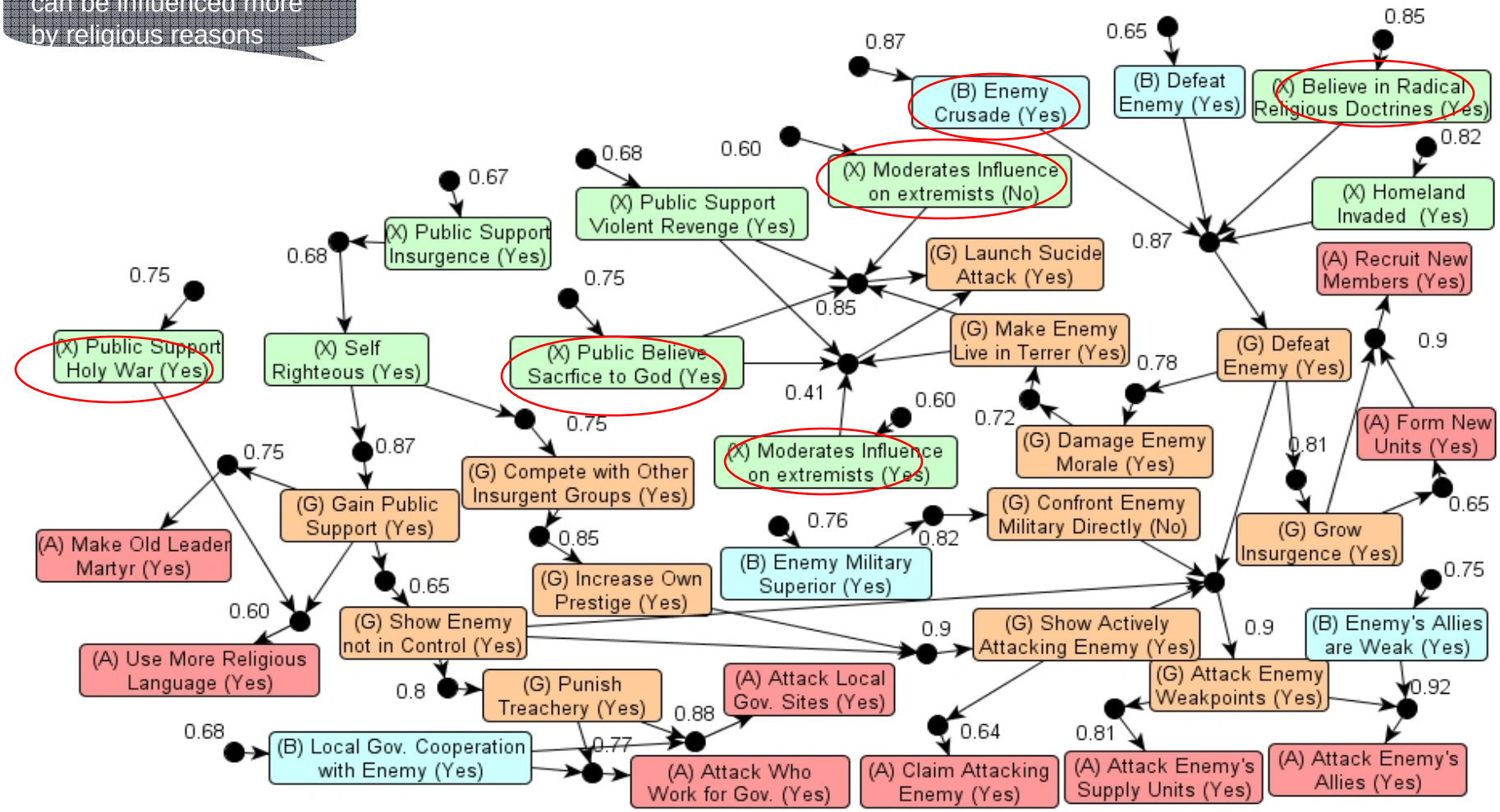
Groups of adversaries and neutrals being driven by DAGA –
reacting to coalition actions based on the current game state and their goals,
internal beliefs, external beliefs, and actions.


0.8
0.2
1.0
0.3
0.7
(G) Gain control of Kadhimiya
Yes
(G) Gain control of Kadhimiya
No
(B) Recent terror attack in Kadhimiya
True
(B) Recent terror attack in Kadhimiya
False
0.96
0.999
1.0
0.7
0.3
0.78
0.11
0.22
0.89
No
(G) Increase influence in Kadhimiya
Yes
(G) Increase influence in Kadhimiya
No
(G) Reduce terrorist activities in Kadhimiya
Yes
(G) Reduce terrorist activities in Kadhimiya
No
(G) Decrease Kadhimiya support for terrorists
Yes
(G) Decrease Kadhimiya support for terrorists
No
0.92
0.64
0.08
0.36
0.8
0.12
0.2
0.88
0.82
0.66
0.91
0.34
0.92
(A) Destroy terrorists in/near Kadhimiya
Yes
(A) Destroy terrorists in/near Kadhimiya
No
(A) Patrol Kadhimiya
Yes
(A) Patrol Kadhimiya
No
0.92
0.08
1.0
0.23
0.77

A Network for Religious based Insurgent Group

(Santos et al. 2007)

Religious insurgent group can be influenced more by religious reasons



Powered by yFiles

Conclusion

- Understanding the intentions of our adversaries is crucial to mitigating threats to our control systems
- Adversarial modeling is the key to obtaining this understanding

- Special thanks to J.T. Wilkinson

Select Control System References

- Adaptive Coordinated Control of Intelligent Multi-Agent Teams, *MURI funded by Army Research Office*, *http://kumar.cis.upenn.edu/acclimate/*, *2002-2005*
- Chandrasekharan, P., C., Robust Control of Linear Dynamical Systems, Academic Press, 1996
- Control System Security Program (CSSP), Department of Homeland Security, http://www.us-cert.gov/control_systems/
- Cooperative Networked Control of Dynamical Peer-to-Peer Vehicle Systems, Sponsored by DDR&E and DARPA/AFOSR, http://legend.me.uiuc.edu/~dullerud/Master/Projects/Project_0.html, 2002-2005
- Franklin, G. F., Emami-Naeini, A., and Powell, J. D. Feedback Control of Dynamic Systems. 3rd. Addison-Wesley Longman Publishing Co., Inc., 1993
- Kalman, R.E. "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering* **82** (1): 35–45, 1960
- Proakis, JG, Anolakis, DG, "**Digital signal processing**: **principles**, **algorithms**, and **applications",** Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1996
- Sastry S. and Bodson M., Adaptive Control: Stability, Convergence, and Robustness, Prentice-Hall, 1989-1994

Select References for Adversarial Modeling

- Santos, Eugene, Jr., Li, Deqing, and Yuan, Xiuqing, "On Deception Detection in Multi-Agent Systems and Deception Intent," *Proceedings of the SPIE: Defense & Security Symposium*, Vol. 6965, Orlando, FL, 2008.
- Santos, Eugene, Jr., McQueary, Bruce, and Krause, Lee, "Modeling Adversarial Intent for Interactive Simulation and Gaming: The Fused Intent System," *Proceedings of the SPIE: Defense & Security Symposium*, Vol. 6965, Orlando, FL, 2008.
- Santos, Eugene, Jr., Zhao, Qunhua, Pratto, Felicia, Pearson, Adam R., McQueary, Bruce, Breeden, Andy, and Krause, Lee, "Modeling Multiple Communities of Interest for Interactive Simulation and Gaming: The Dynamic Adversarial Gaming Algorithm Project," *Proceedings of the SPIE: Defense & Security Symposium*, Vol: 6564, Orlando, FL, 2007.
- Santos, Eugene, Jr. and Zhao, Qunhua, "Adversarial Models for Opponent Intent Inferencing," in *Adversarial Reasoning: Computational Approaches to Reading the Opponents Mind* (Eds. A. Kott and W. McEneaney), pp 1-22, Chapman & Hall/CRC Computer and Information Science Series, Chapman & Hall/CRC:Boca Raton, 2006.
- Santos, Eugene, Jr., Zhao, Qunhua, Johnson, Gregory, Nguyen, Hien, and Thompson, Paul, "A Cognitive Framework For Information Gathering with Deception Detection For Intelligence Analysis," *Proceedings of the 2005 International Conference on Intelligence Analysis*, McClean, VA, 2005.
- Lehman, Lynn A., Krause, Lee S., Gilmour, Duane A., Santos, Eugene, Jr., and Zhao, Qunhua "Intent Driven Adversarial Modeling," *Proceedings of the Tenth International Command and Control Research and Technology Symposium: The Future of C2*, McLean, VA, 2005.
- Santos, Eugene, Jr. and Johnson, Gregory, "Toward Detecting Deception in Intelligent Systems," *Proceedings of the SPIE: Defense & Security Symposium*, Vol. 5423, 131-140, Orlando, FL 2004.
- Santos, Eugene, Jr. and Negri, Allesandro, "Constructing Adversarial Models for Threat Intent Prediction and Inferencing," *Proceedings of the SPIE Defense & Security Symposium*, Vol. 5423, 77-88, Orlando, FL 2004.
- McQueary, Bruce, Krause, Lee, Santos, Eugene, Jr., Wang, Hua, and Zhao, Qunhua, "Analysis and Visualization of Uncertainty in the Battlespace," *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004)*, 782-784, Boca Raton, FL, 2004.
- Santos, Eugene, Jr., "A Cognitive Architecture for Adversary Intent Inferencing: Knowledge Structure and Computation," *Proceedings of the SPIE: 17th Annual International Symposium on Aerospace/Defense Sensing and Controls: AeroSense 2003*, Vol. 5091, 182-193, Orlando, FL, 2003.
- Surman, Joshua, Hillman, Robert, and Santos, Eugene, Jr., "Adversarial Inferencing for Generating Dynamic Adversary Behavior," *Proceedings of the SPIE: 17th Annual International Symposium on Aerospace/Defense Sensing and Controls: AeroSense 2003*, Vol. 5091, 194-201, Orlando, FL, 2003.
- Brown, Scott M., Santos, Eugene, Jr., and Bell, Benjamin, "Knowledge Acquisition for Adversary Course of Action Prediction Models," *Proc of the AAAI 2002 Fall Symposium on Intent Inference for Users, Teams, and Adversaries*, Boston, MA, 2002.
- Bell, Benjamin, Santos, Eugene, Jr., and Brown, Scott M., "Making Adversary Decision Modeling Tractable with Intent Inference and Information Fusion," *Proceedings of the 11th Conference on Computer Generated Forces and Behavioral Representation*, 535-542, Orlando, FL, 2002.

Select References for User Modeling

- Nguyen, Hien and Santos, Eugene, Jr., "An Evaluation of the Accuracy of Capturing User Intent for Information Retrieval," *Proc. Of the Intl Conf on AI*, Las Vegas, NV, 2007.
- Nguyen, Hien and Santos, Eugene, Jr., "Effects of Prior Knowledge on the Effectiveness of a Hybrid User Model for Information Retrieval," *Proceedings of the SPIE: Defense & Security Symposium*, Vol: 6536, Orlando, FL, 2007.
- Nguyen, Hien, Santos, Eugene, Jr., Smith, Nathan, and Schuett, Aaron, "Hybrid User Model for Information Retrieval," *Proceedings of the AAAI Workshop on Modeling Others from Observations (MOO-06)*, 61-76, Boston, MA, 2006.
- Santos, Eugene, Jr., Zhao, Qunhua, Nguyen, Hien, and Wang, Hua, "Impacts of User Modeling on Personalization of Information Retrieval: An Evaluation with Human Intelligence Analysts," in S. Weibelzahl, A. Paramythis, and J. Masthoff (Eds.), *Proceedings of the Fourth Workshop on the Evaluation of Adaptive Systems* (held in conjunction with the 10th International Conference on User Modeling (UM'05)), 27-36, Edinburgh, UK, 2005.
- Nguyen, Hien, Santos, Eugene, Jr., Zhao, Qunhua, and Wang, Hua, "Capturing User Intent for Information Retrieval," *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society (HFES 2004)*, 371-375, New Orleans, LA, 2004.
- Nguyen, Hien, Santos, Eugene, Jr., Zhao, Qunhua, and Lee, Chester, "Evaluation of Effects on Retrieval Performance for an Adaptive User Model," *Adaptive Hypermedia 2004: Workshop Proceedings - Part I*, 193-202, Eindhoven, The Netherlands, 2004.
- McQueary, Bruce, Krause, Lee, Santos, Eugene, Jr., Wang, Hua, and Zhao, Qunhua, "Analysis and Visualization of Uncertainty in the Battlespace," *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004)*, 782-784, Boca Raton, FL, 2004.
- Santos, Eugene, Jr., Nguyen, Hien, Zhao, Qunhua, and Wang, Hua, "User Modelling for Intent Prediction in Information Analysis," *Proceedings of the 47th Annual Meeting for the Human Factors and Ergonomics Society (HFES-03)*, 1034-1038, Denver, CO, 2003.
- Santos, Eugene, Jr., Nguyen, Hien, Zhao, Qunhua, and Pukinskis, Erik, "Empirical Evaluation of Adaptive User Modeling in a Medical Information Retrieval Application," *Lecture Notes in Artificial Intelligence 2702: User Modeling 2003* (Eds. P. Brusilovsky, A. Corbett, and F. de Rosis), 292-296, Springer, 2003.
- Santos, Eugene, Jr., Nguyen, Hien, and Brown, Scott M., "Kavanah: An Active User Interface Information Retrieval Application," *Proceedings of the 2nd Asia-Pacific Conference on Intelligent Agent Technology*, 412-423, 2001.
- Santos, Eugene, Jr., Brown, Scott M., Lejter, Moises, Ngai, Grace, Banks, Sheila B., and Stytz, Martin R., "Dynamic User Model Construction with Bayesian Networks for Intelligent Information Queries," *Proceedings of the 12th International FLAIRS Conference*, 3-7, Orlando, FL, 1999.
- Brown, Scott M., Santos, Eugene, Jr., and Banks, Sheila B., "Utility theory-based user models for intelligent interface agents," *Lecture Notes in Artificial Intelligence 1418: Advances in Artificial Intelligence – AI '98*, 378-392, Springer-Verlag, 1998.
- Brown, Scott M., Santos, Eugene, Jr., Banks, Sheila B., and Oxley, Mark, "Using Explicit Requirements and Metrics for Interface Agent User Model Correction," *Proceedings of the Second International Conference on Autonomous Agents*, 1-7, Minneapolis/ St. Paul, MN, 1998.
- Brown, Scott M., Harrington, Robert A., Santos, Eugene, Jr., and Banks, Sheila B., "User Models, Intelligent Interface Agents and Expert Systems," *Proceedings of the Sixth International Conference on User Modeling Workshop on Embedding User Models in Intelligent Applications*, 12-16, Sardinia, Italy, 1997.
- Santos, Eugene Jr., Nguyen, Hien , Yu, Fei, Kim, Keumjoo, Li, Deqing, Wilkinson, John, Olson, Adam , and Russell, Jacob, "Intent-driven insider threat detection in intelligence analyses," to appear Proceedings of IEEE/WIC/ACM International Conference on Intelligent Agent Technology, 2008.

Select References for Bayesian Knowledge Bases

- Santos, Eugene, Jr. and Dinh, Hang T., "Automatic Knowledge Validation for Bayesian Knowledge Bases," *Data and Knowledge Engineering* **64**, 218-241, 2008.
- Rosen, Tzachi, Shimony, Solomon Eyal, and Santos, Eugene, Jr., "Reasoning with BKBs – Algorithms and Complexity," *Annals of Mathematics and Artificial Intelligence* **40(3-4)**, 403-425, 2004.
- Santos, Eugene, Jr., Santos, Eugene S., and Shimony, Solomon Eyal, "Implicitly Preserving Semantics During Incremental Knowledge Base Acquisition Under Uncertainty," *International Journal of Approximate Reasoning* **33(1)**, 71-94, 2003.
- Santos, Eugene, Jr., "Verification and Validation of Knowledge-Bases Under Uncertainty," *Data and Knowledge Engineering* **37**, 307-329, 2001.
- Johnson, Gregory and Santos, Eugene, Jr., "Generalizing Knowledge Representation Rules for Uncertain Knowledge," *Proceedings of the 13th International FLAIRS Conference*, 186-190, Orlando, FL, 2000.
- Santos, Eugene, Jr. and Young, Joel D., "Probabilistic Temporal Networks: A unified framework for reasoning with Time and Uncertainty," *International Journal of Approximate Reasoning* **20**, 263-291, 1999.
- Santos, Eugene, Jr. and Santos, Eugene S., "A Framework for Building Knowledge-Bases Under Uncertainty," *Journal of Experimental and Theoretical Artificial Intelligence* **11**, 265-286, 1999.
- Santos, Eugene, Jr., and Shimony, Solomon Eyal, "Deterministic Approximation of Marginal Probabilities in Bayes Nets," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans* **28(4)**, 377-393, 1998.
- Santos, Eugene, Jr., "On Linear Potential Functions for Approximating Bayesian Computations," *Journal of the ACM* **43(3)**, 399-430, 1996.
- Santos, Eugene, Jr., "On the Generation of Alternative Explanations with Implications for Belief Revision," *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, 337-347, Los Angeles, CA, 1991.
- Santos, Eugene S., and Santos, Eugene, Jr., "Reasoning with Uncertainty in a Knowledge-Based System," *Proceedings of the Seventeenth International Symposium on Multiple-Valued Logic*, 75-81, Boston, MA, 1987.

BKBs vs BNs (1/2)

Probabilistic models exhibiting significant local structure are common. In such models, explicit representation of that structure as done in BKBs, is advantageous, as the resulting representation is much more compact than the full table representation of the conditional probability tables (CPTs) in a BN. For example, consider the following setting: X, a binary variable, is known to be true if any of the variables Y_i is true, for $1 \leq i \leq n$, and X is false with probability p otherwise. The global structure here is that X depends on all the Y_i and in a BN one might represent this with a set of arcs $\{(Y_i, X) \mid 1 \leq i \leq n\}$. The representation of the distribution in the "standard" form

BKBs vs BNs (2/2)

of a CPT would require $O(2^n)$ entries. However, the (partially) given distribution also exhibits "local" structure, as when Y_i is known to be true for some i, X no longer depends on the value of Y_j for $j \neq i$. The size of the representation of the conditional probabilities in terms of rules is only $O(n)$. Although work has been done on representing local structure using other methods, such as local decision trees and default tables [4], rules have significant advantages in size of the representation, as well as their better explainability. For example, contrasting rules with decision trees as a representation of local structure, every decision tree is compactly representable as a set of rules, while the reverse is not necessarily true - the decision tree may be exponentially larger that the set of rules [1]. Although rule-based systems for representing an exact distribution exist (e.g. [15]), these systems are a (compact) notational variant of Bayes networks, and are thus less flexible than BKBs, as they do not allow for incompleteness or cyclicity.