Question Answering in the InfoSphere: Semantic Interoperability and Lexicon Development

Paul Thompson
and
Steven Lulich

Institute for Security Technology Studies
Dartmouth College

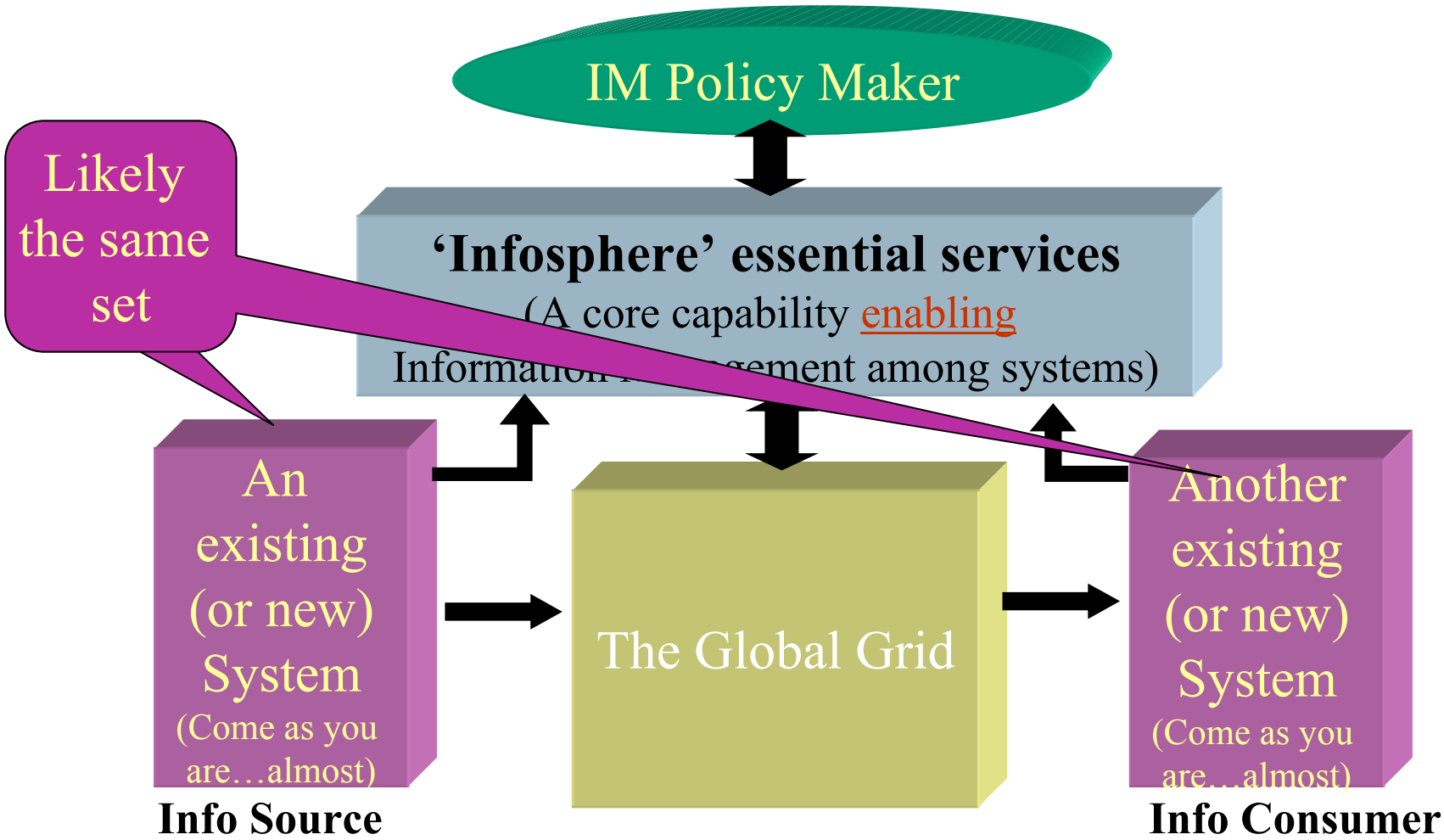
15 May 2002

Outline

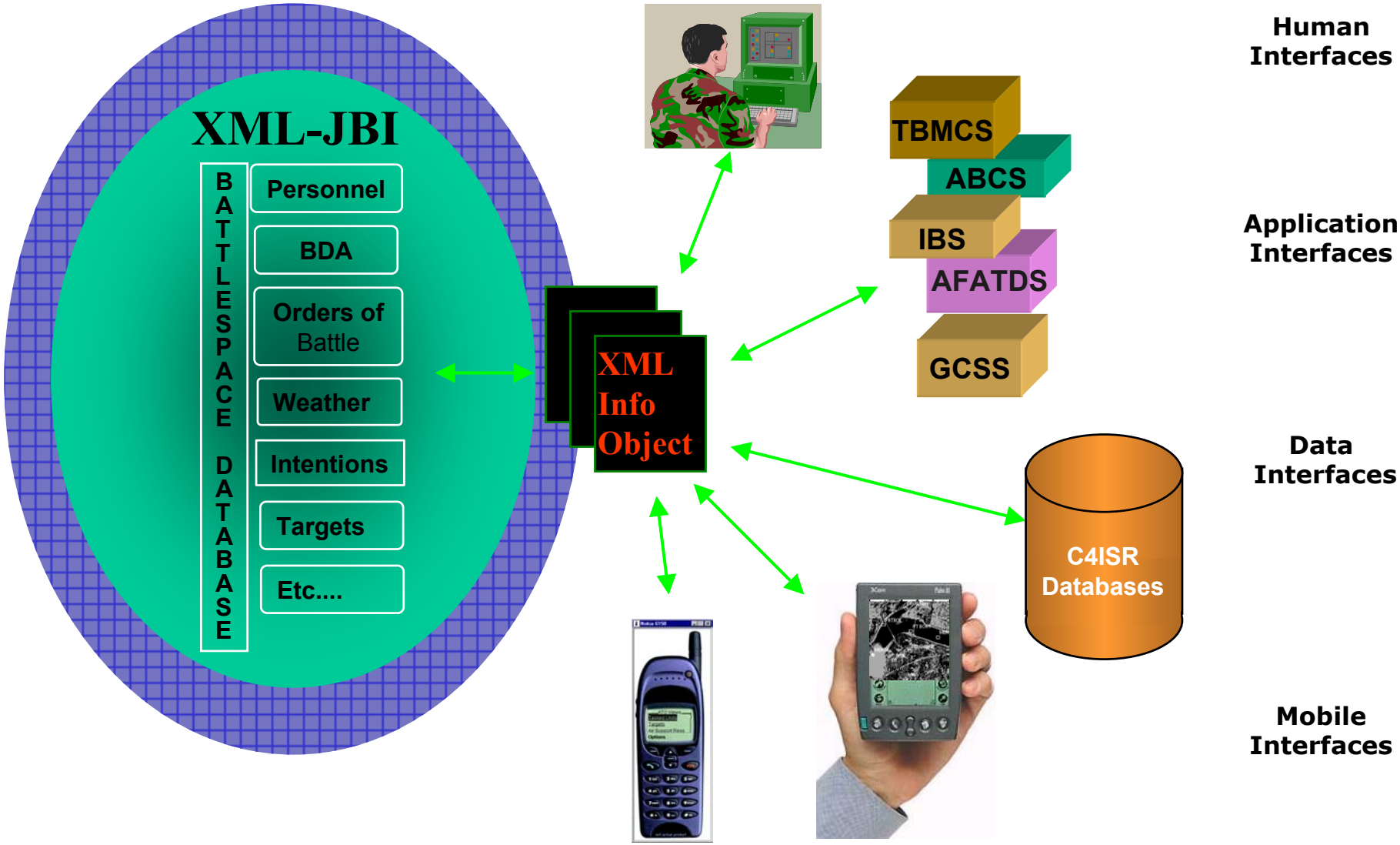
- The Joint Battlespace Infosphcre
- Fuselets
- DARPA IXO Sensor Networks
- Multiple Query Optimization
- Semantic Interoperability and Lexicon Development
- Conclusions

Basic "Infosphere" concept

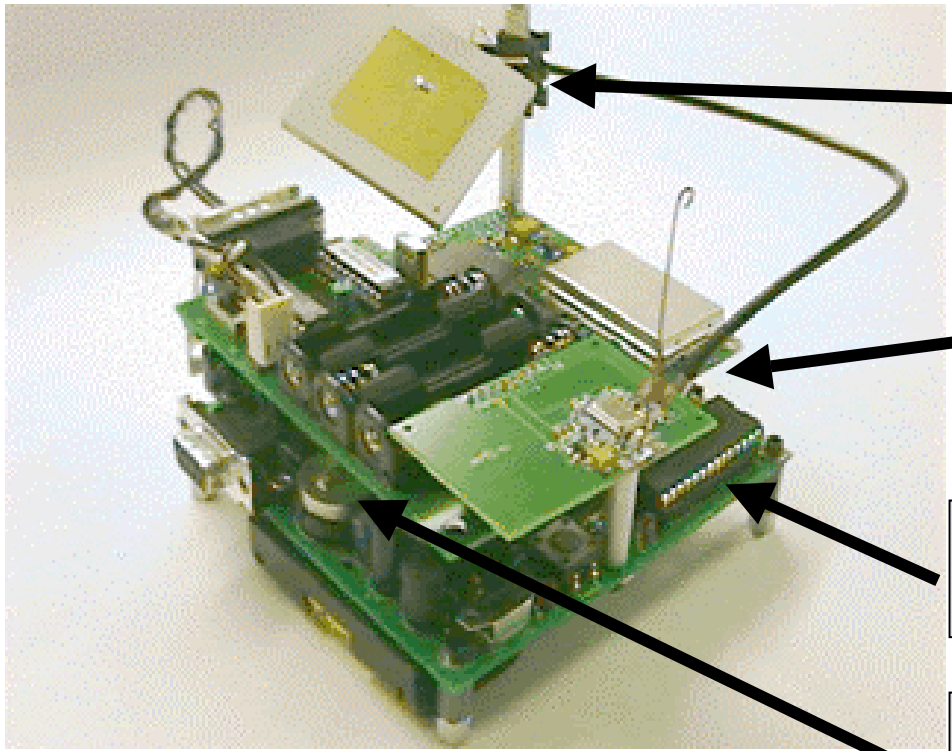
- An information broker will offer essential information exchange services



Information Interoperability using XML



Dartmouth Physical Sensor (Version 1)



GPS Location Sensor (Motorola)

RF Technologies Wireless Networking

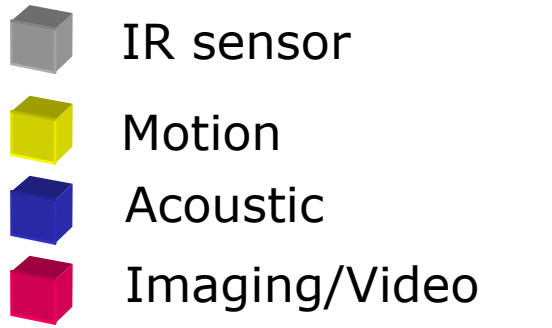
Processor (Intel 8051)

Dallas Semiconductor I-Buttons

DURIP funded 100+ such devices

Other efforts: Crossbow, UC Berkeley, UCLA,....

Challenge Problem

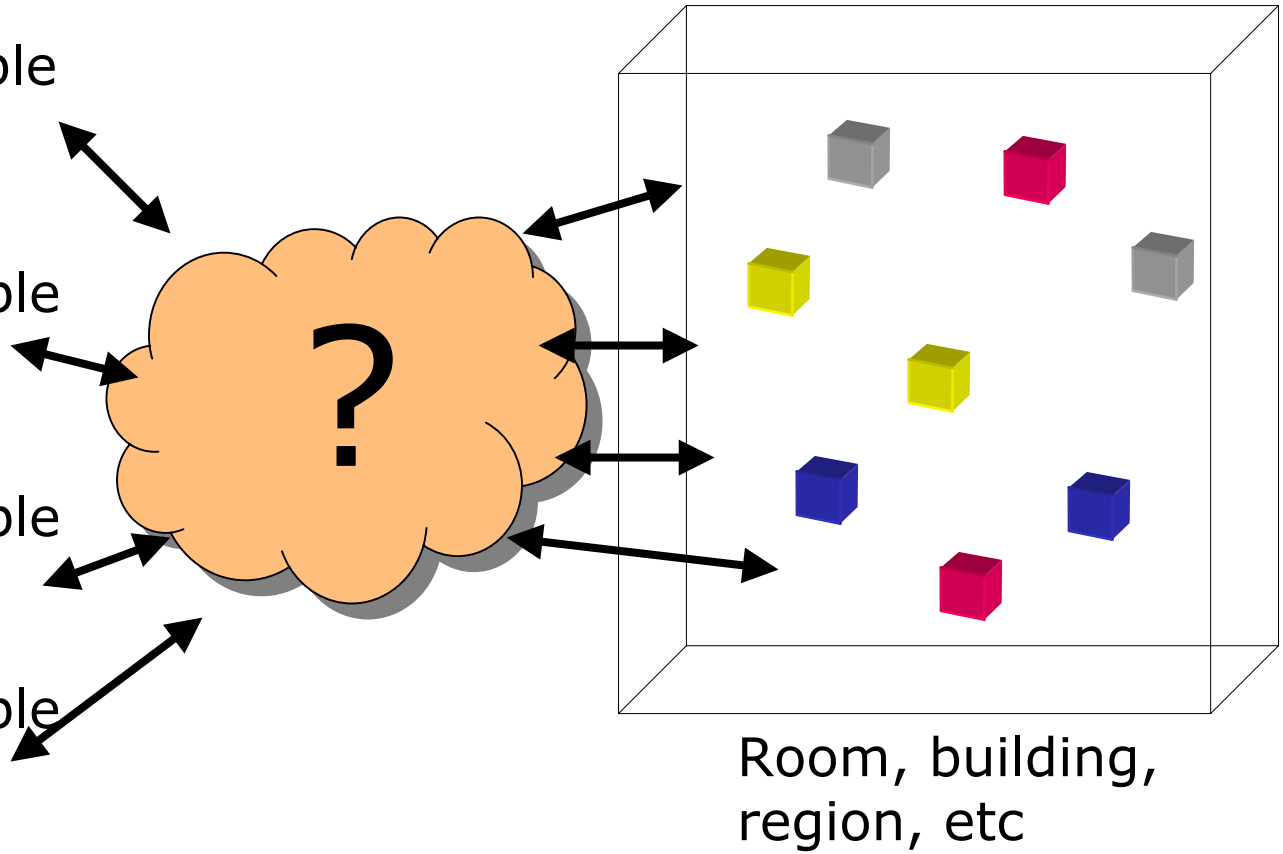


Q1: "How many people are in the room?"

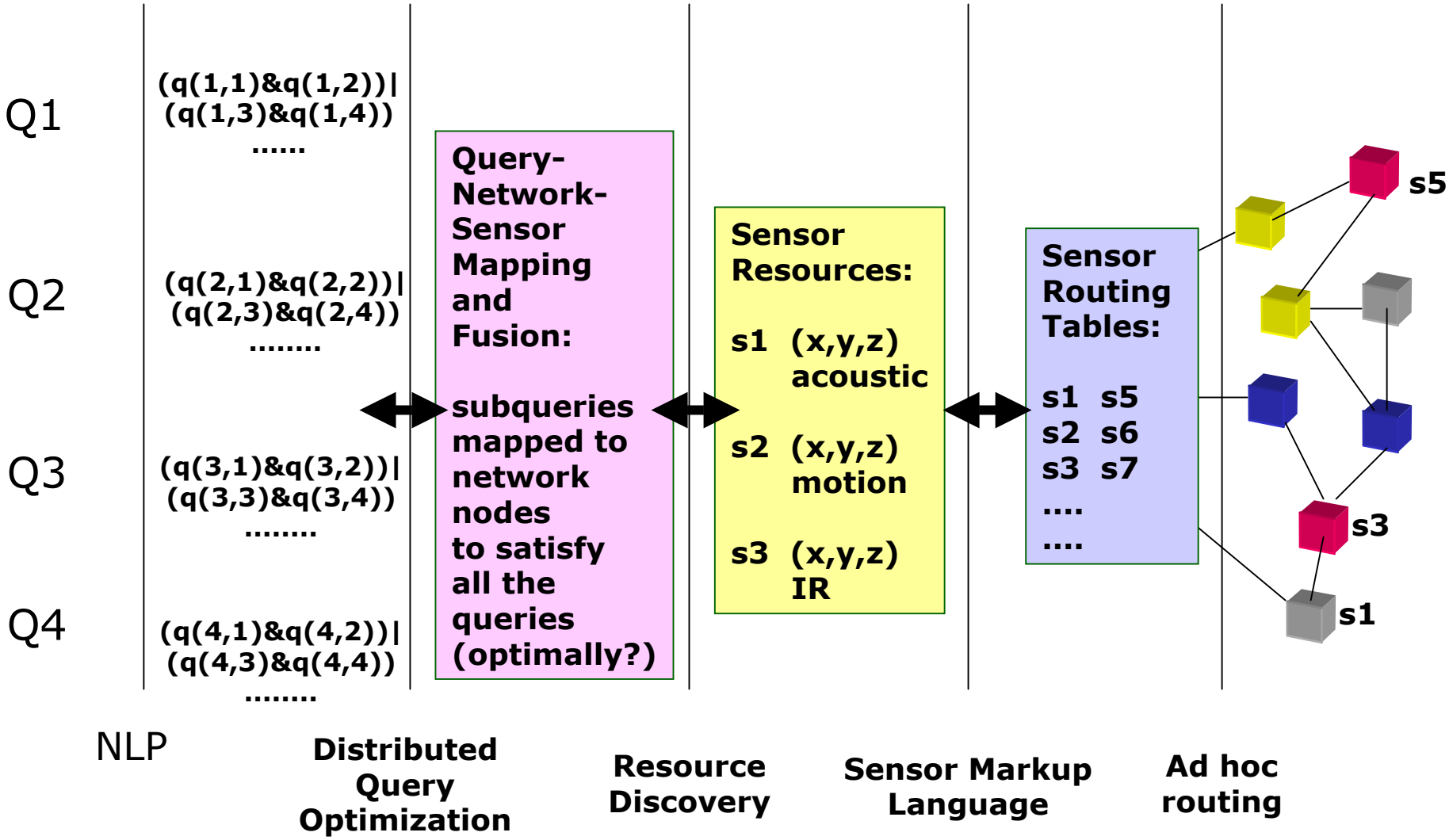
Q2: "How many people are in the room?"

Q3: "How many people are talking?"

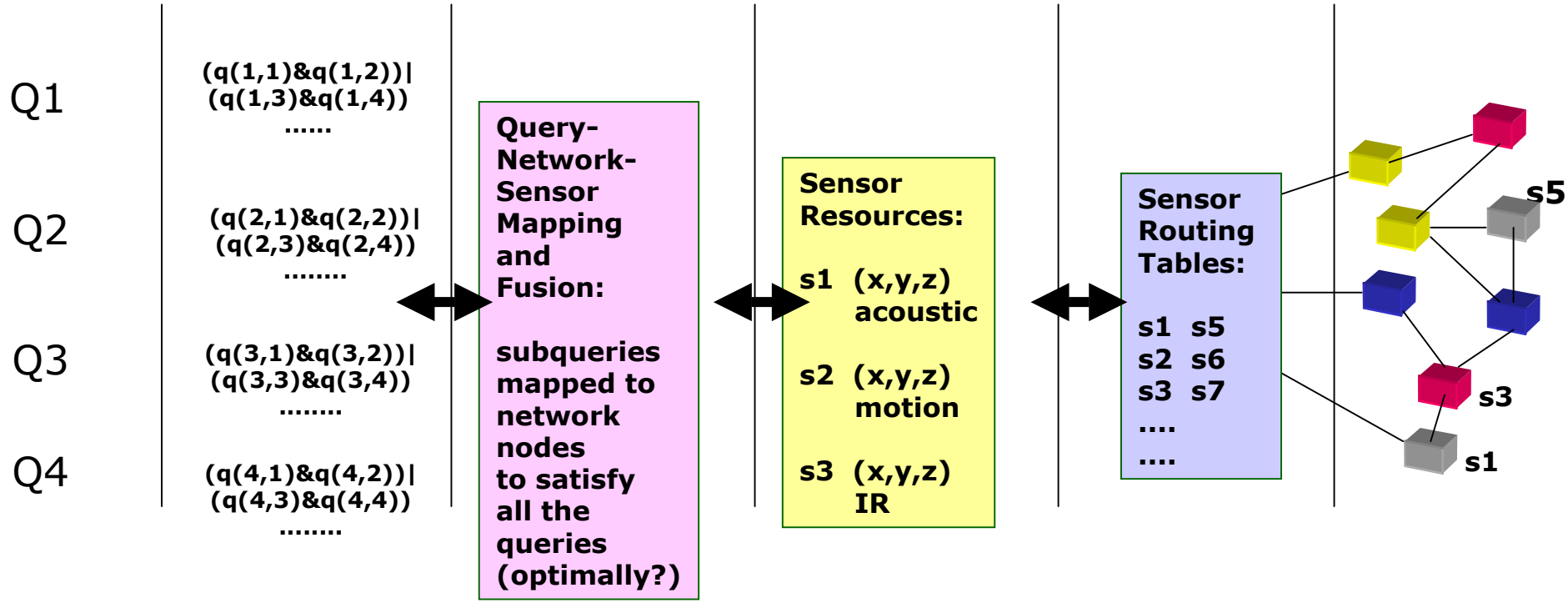
Q4: "How many people are listening?"



Ingredients ("Sensor Net OSI Layers?")



Ingredients - Score Card



NLP	Distributed Query Optimization	Resource Discovery	Sensor Markup Language	Ad hoc routing
Hard	NP Hard, approximate?	Scalability in time and numbers?	Semantics, consistency, dynamics?	Scalability in time and numbers?

Natural Language Processing Level

- "Question and Answer" systems
  - TREC Question Answering Track, eg
  - "Who won the women's downhill gold medal?"
- "Message Understanding"
  - Bounded domains - military intelligence, logistics, etc
  - message retrieval using information extraction
- Inference capability

Semantic Interoperability and Lexicon Development

- DARPA hand-built Ontologies and KBs
  - DARPA Agent Markup Language (DAML)
  - Rapid Knowledge Formation (RKF)
- Dynamic Ontology Mapping
  - Learning component
  - Connectivistic Lexicon

Distributed Query Optimization Layer

User 1: How many fuel trucks are there in region x?

Parsed to a boolean search expressed in terms of sensor outputs and primitive operations.

q(u1,si,opj)&q(u1,sk,opn)|....

User 2: Where are the armored personnel carriers in region y? (y intersects x)

q(u2,si2,opj2)&q(u2,sk2,opn2)|....

Implement this in a network of fuselet servers using relationships, redundancies, inclusions, etc to get good performance according to some metric.

Effectiveness and Efficiency of a Mapping

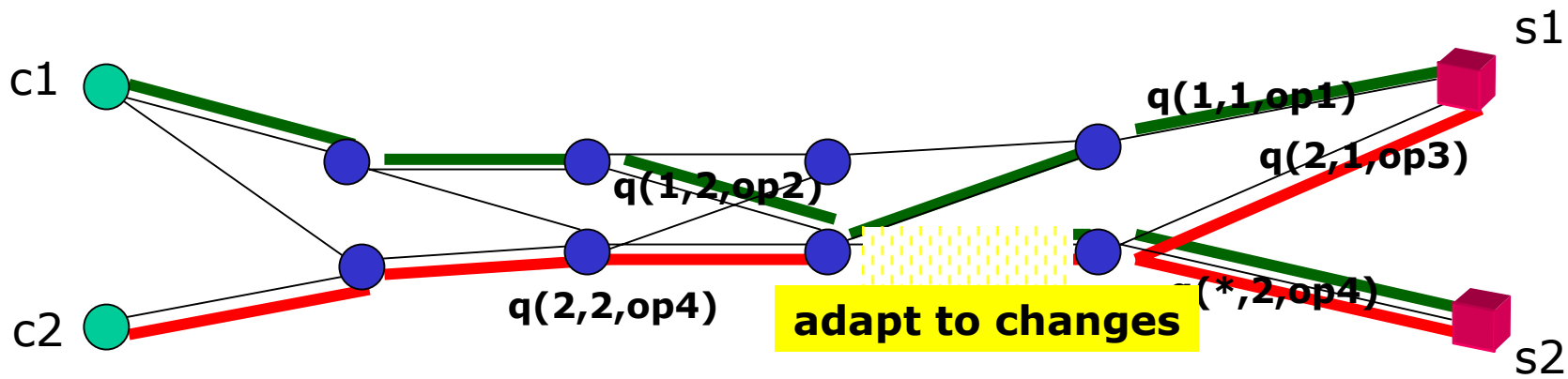
q(1,1,op1) & q(1,2,op2)

q(2,1,op3) | q(2,2,op4)

q(1,2,op2) < q(2,2,op4)

& means downstream flow is smaller

| means downstream flow is the same

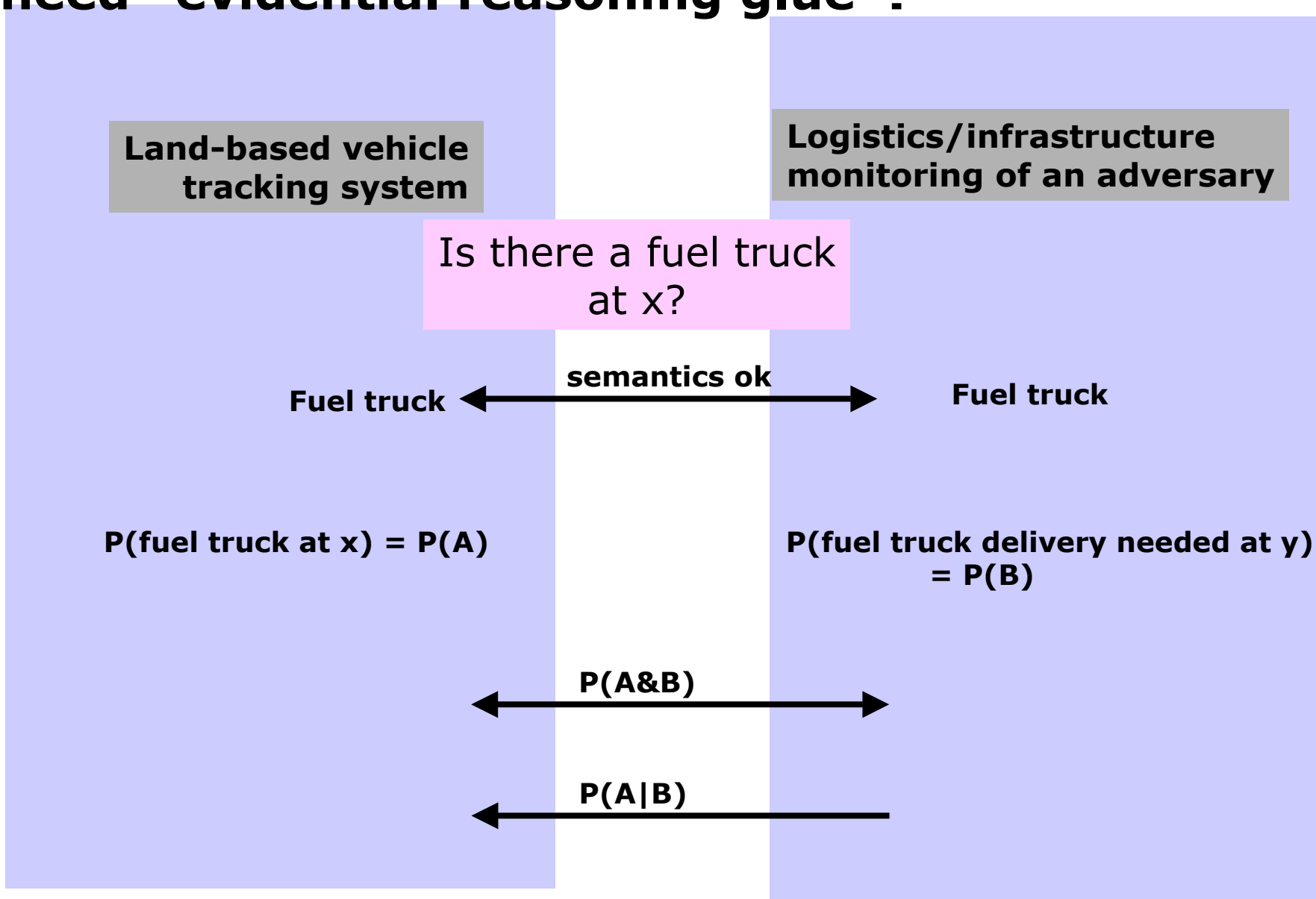


q(client,sensor,operation)

Resource Discovery

- JINI, CORBA, Brokers, Matchmakers, etc
- Distributed, coherency, overheads, etc
- Semantics?
- Indexing?
- Scalability of distributed, dynamic implementation

Sensor Markup Language (XML see DARPA DAML) - need "evidential reasoning glue"?



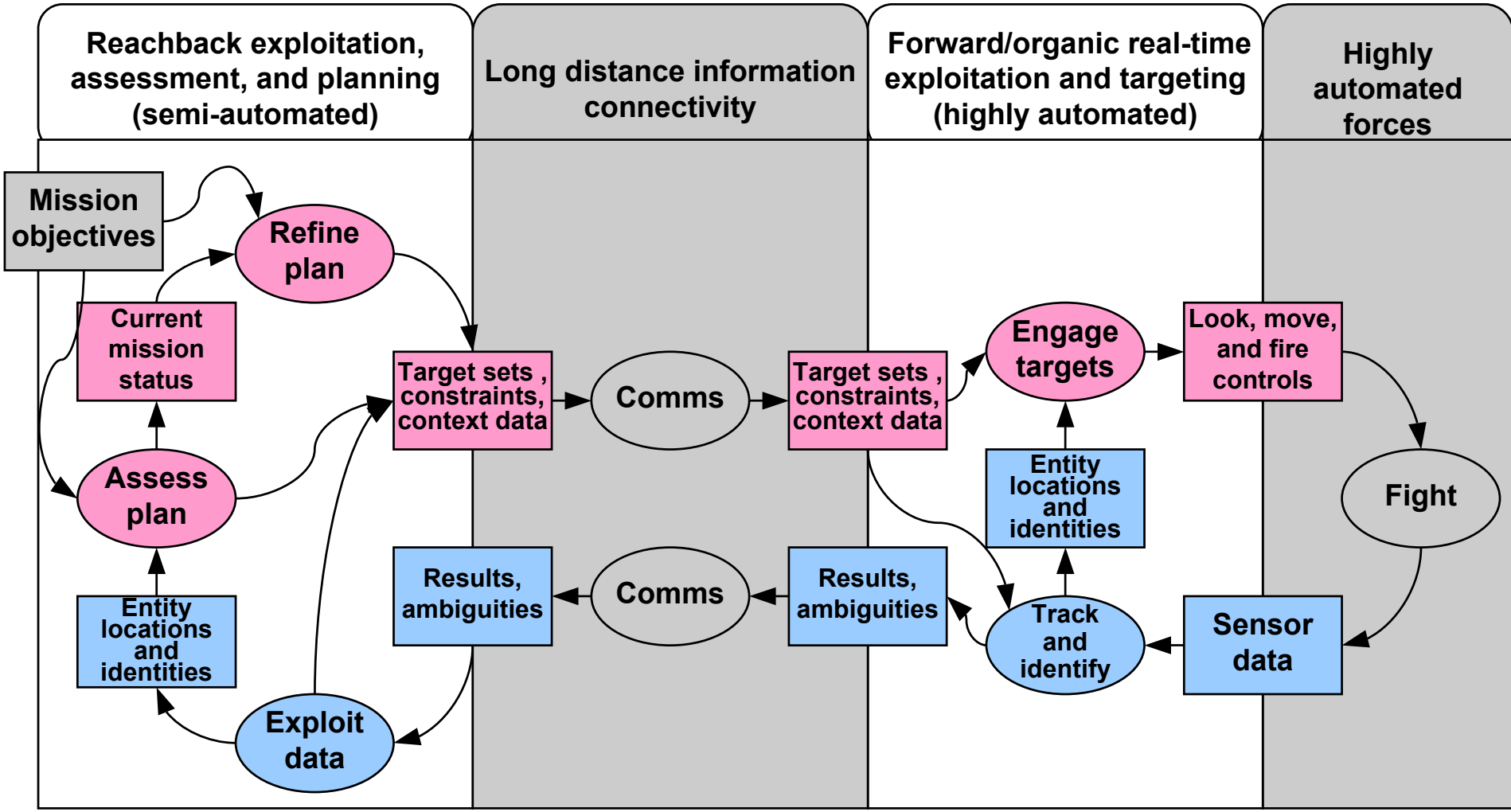
Ad hoc routing in dynamic environments

- nodes dynamic
- requirements dynamic
- tradeoff between management overhead and usable BW/latency?
- sensor networks: bi-directional vs unidirectional links?
- lots to do still

C4ISR for Tactical Air-Land Combat

System Concept

Nested, Closed-loop Control of Highly Automated Forces



Conclusions (cont.)

- Question answering from JBI and sensor networks
  - Different issues from ARDA / TREC
  - Can build on
    - ARDA / TREC
    - NLP interfaces to RDBMS and Knowledge Bases
    - Data mining from stream data
- Semantic Interoperability is crucial