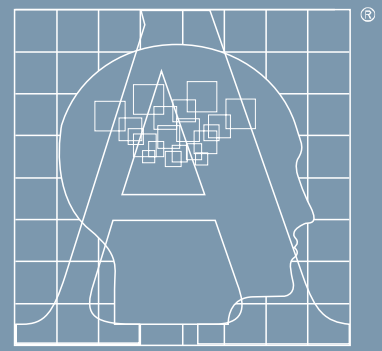


Identifying the Enemy – Part I: Automated Network Identification Model

Georgiy Levchuk, Yuri Levchuk, Elliot Entin
Aptima Inc.

Feili Yu, Haiying Tu, Krishna Pattipati
University of Connecticut

Presented at CCRTS-2007
Date: 6/19/2007

www.aptima.com
Woburn, MA ▪ Washington, DC

Outline

- The problem
- DARPA seedling project
- Proposed solution: NetSTAR
- NetSTAR model
- NetSTAR performance analysis

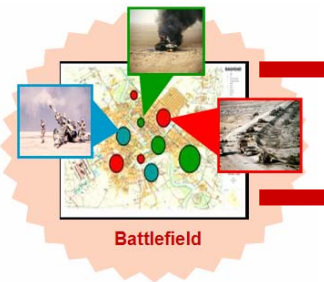
APTIMA HUMAN-CENTERED ENGINEERING

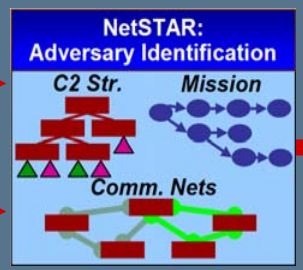
The Problem

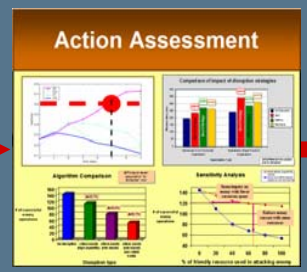
Organization Identification as Part of Larger Problem

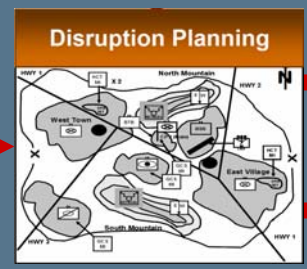
Semi-automated Counteractions Plans Development System

INPUT









PRODUCT

CMDR & planning staff

- *Execute attacks against enemy*
- *Gather intel*

- •comm. intercepts
- •events
- •actions

predicted target network structures, actor transaction patterns, roles & intent

vulnerabilities & counter-action impact assessment

counteractions, disruptions & probes plan

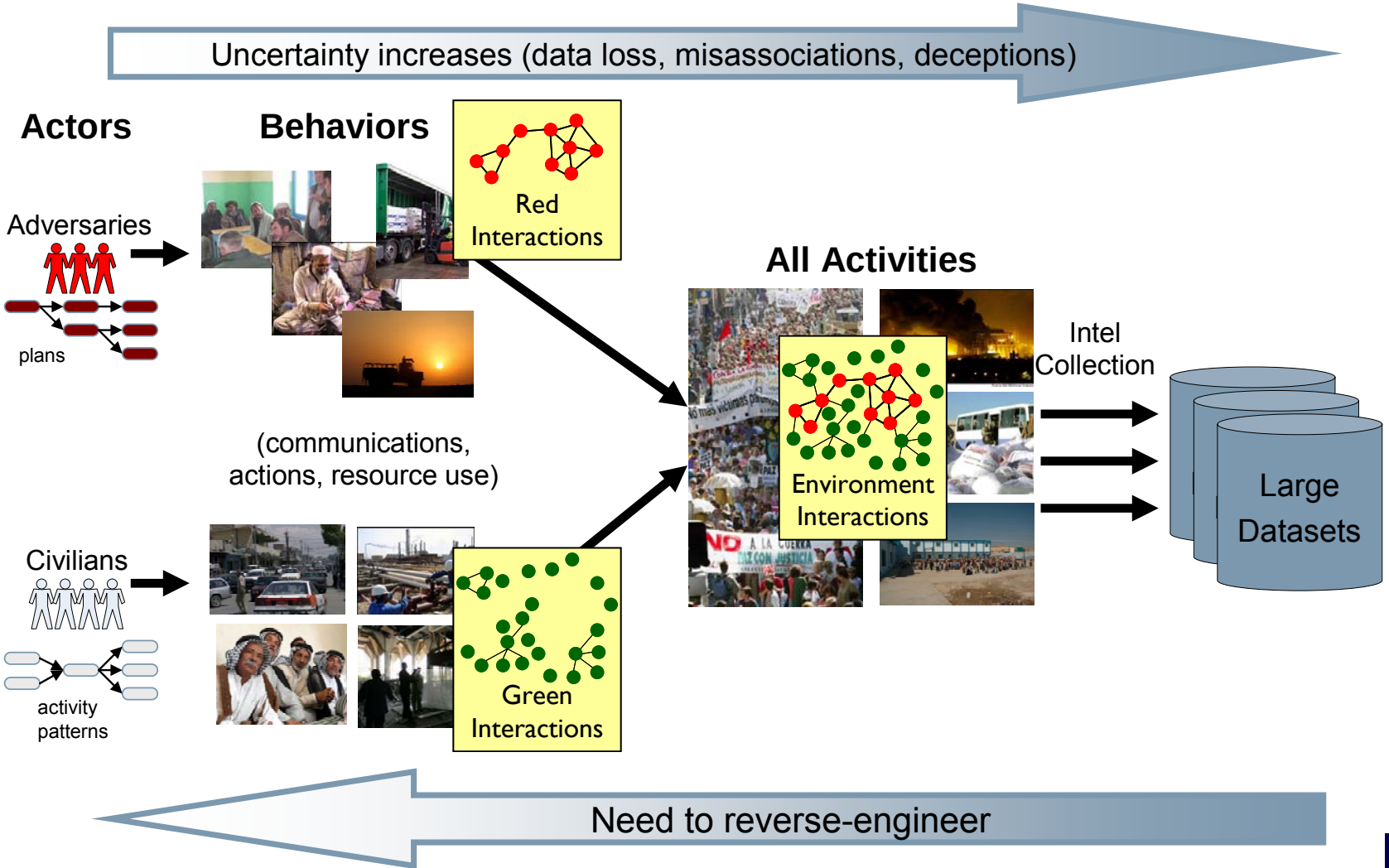
NetSTAR Benefits to Process

- Increase amount of data that can be analyzed and uncertainty/complexity that can be handled
- Speed-up & improve accuracy of threat analysis
- Improve understanding of enemy's vulnerabilities and effectiveness of counteractions
- Improve data collection by targeting most critical info

Outcome: enemy's performance degrades

- incorrect actions
- delayed commands
- missed critical information and engagements

Challenge of Threat Analysis



APTIMA
HUMAN-CENTERED
ENGINEERING

DARPA Seedling Project Focus

Find:

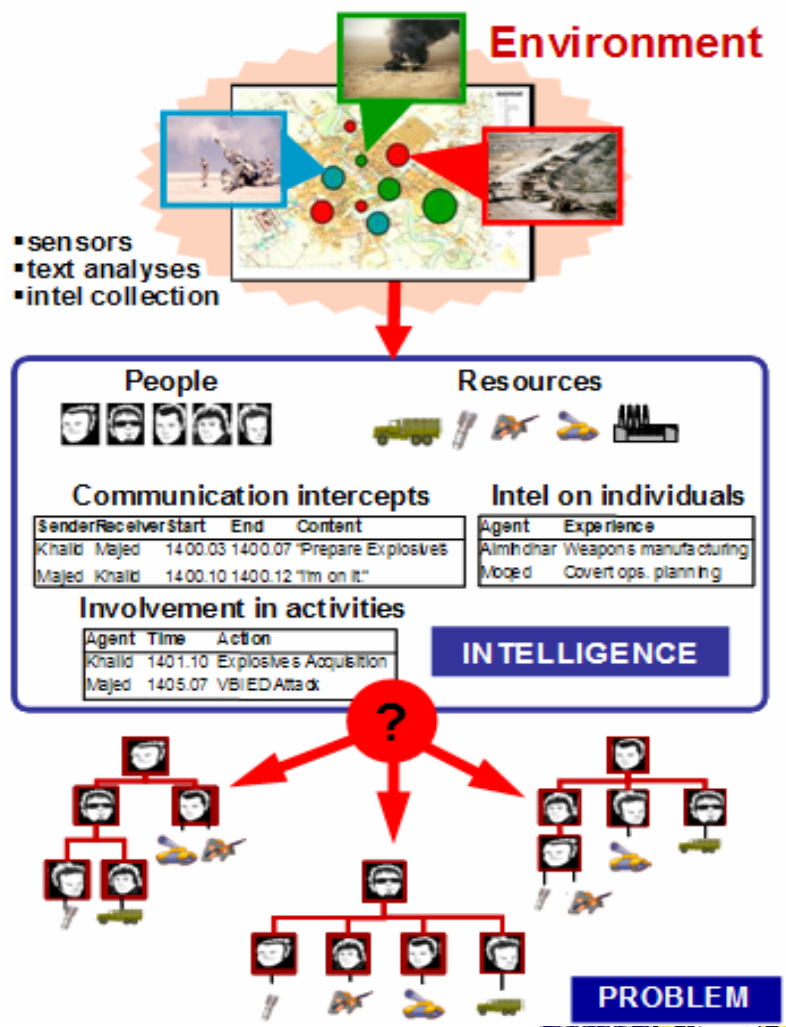
- Enemy **STRUCTURE**
- Enemy **INTENT**
- Enemy **ACTIVITIES**

This will enable you to:

- Find correct RED high-value targets
- Develop effective BLUE COAs/counteractions
- Avoid unintended consequences of BLUE actions

Challenges of manual threat identification

- Enemy adapting – cannot rely on experience only
- Data explosion – high manpower needs, manual approaches would not scale
- Large info gaps & complexity
- Biases in human decisions

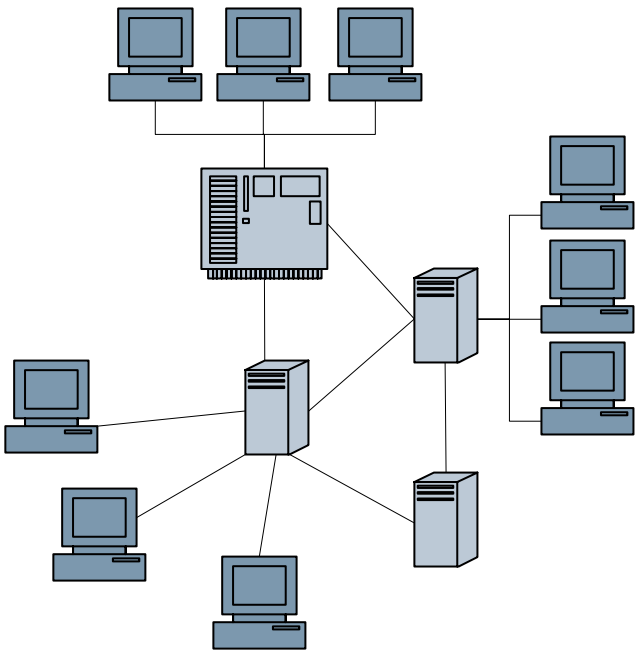


NetSTAR in a Nutshell

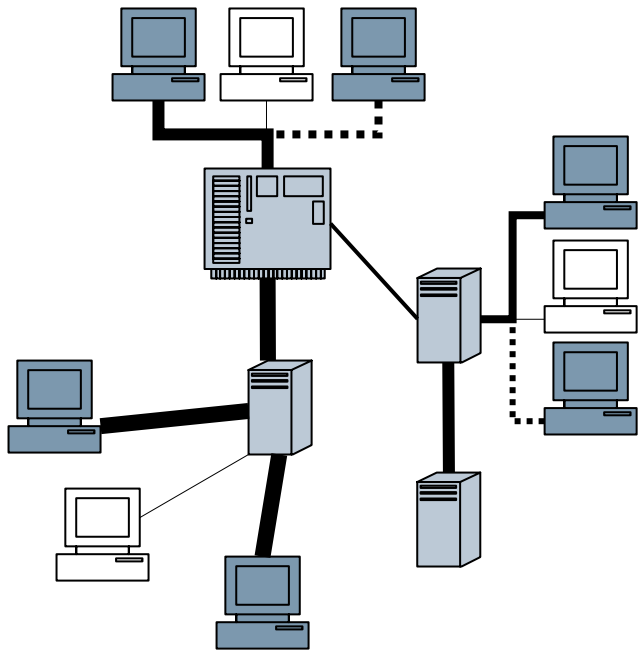
- **What is NetSTAR?**
  - Semi-automated technology to discover transaction patterns and organization network structures from massively noisy data
- **What data does NetSTAR need?**
  - Communication transactions, activities, and actors + Pattern library
- **What makes NetSTAR unique?**
  - Combines organizational science and probabilistic computational models with intelligence analysts' experience
- **What are NetSTAR key benefits for the intelligence analyst?**
  - Reduce the "size of haystack" in search for the needle
  - Allow more time for the analyst to explore relevant information

NetSTAR Idea-1

- **Organization** = infrastructure



- **Interaction pattern** = use of infrastructure



Difference because of what is needed to be done

APTIMA HUMAN-CENTERED ENGINEERING

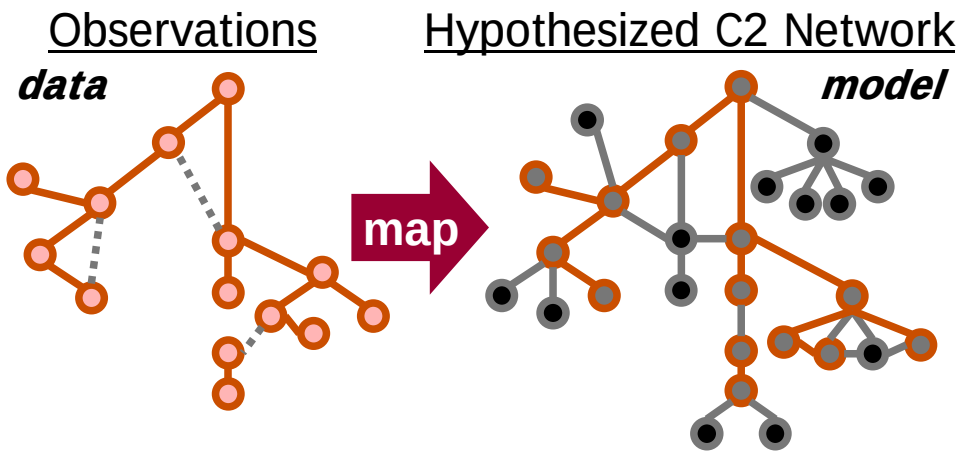
NetSTAR Idea-2

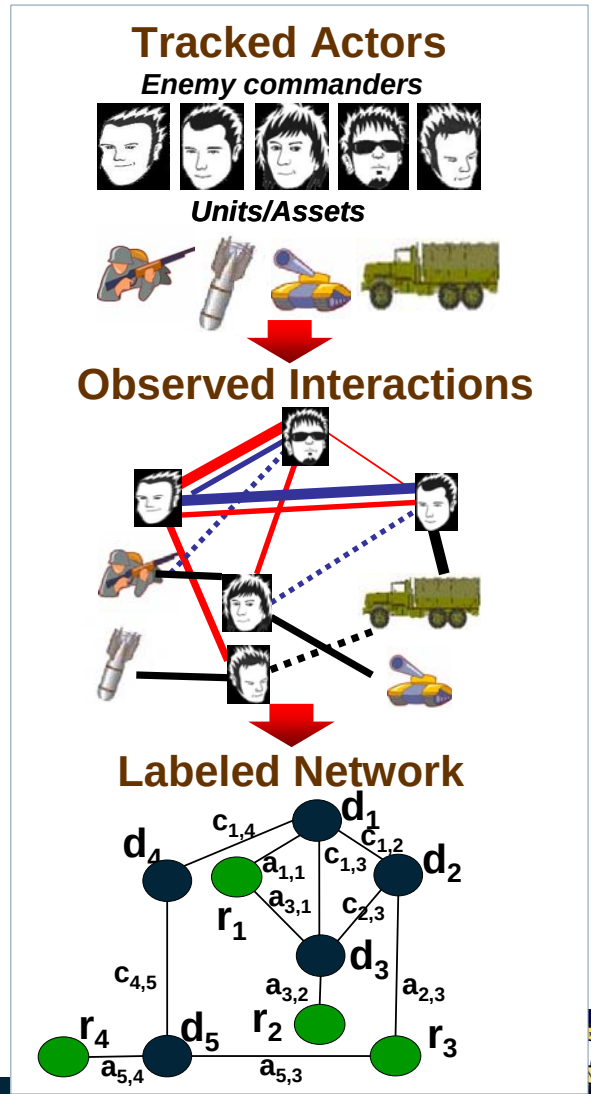
Representation

- C2 organizations can be represented as graphs with labels
  - **Node labels** = actor profiles
  - **Link labels** = type & frequency of interactions

Formalization

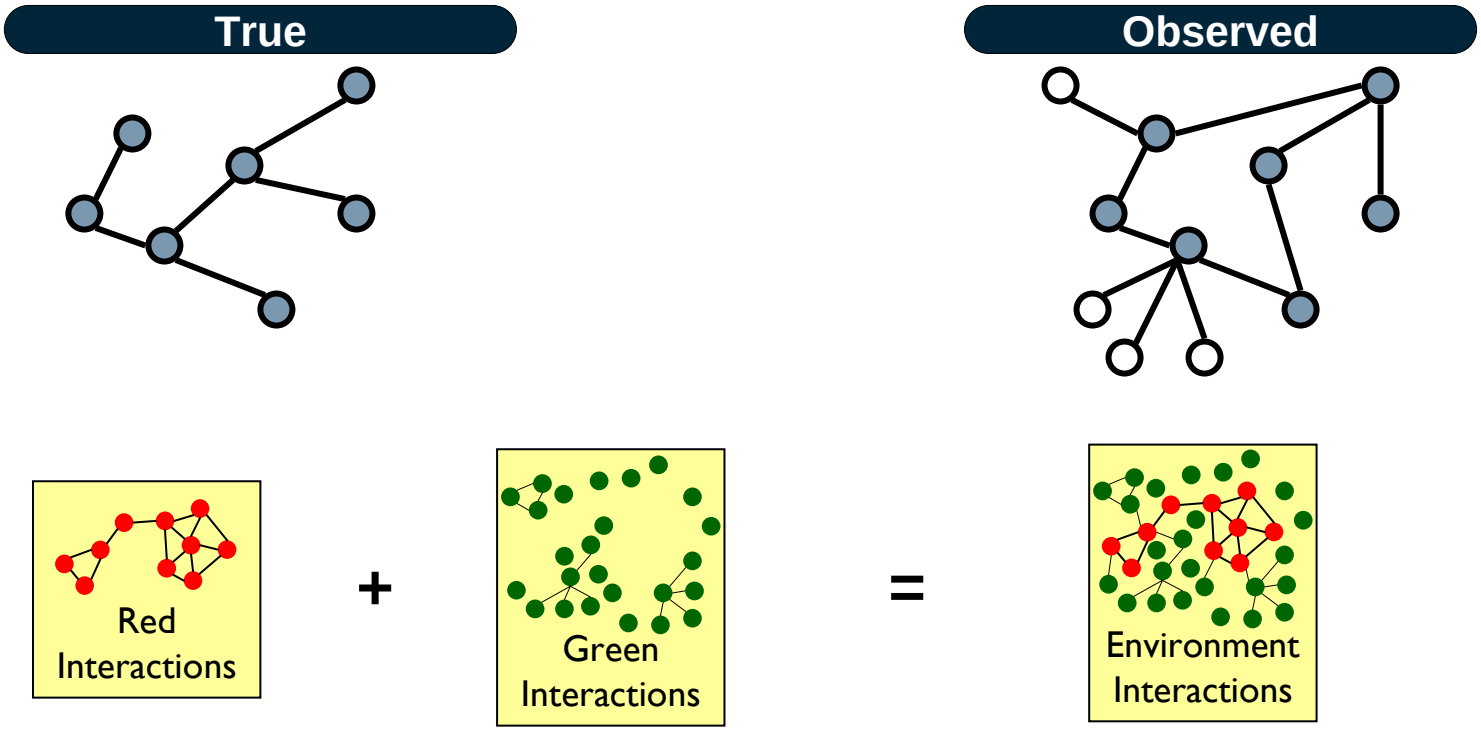
- Find best node-to-node mapping between data & model nets
- Select C2 structure with best map score



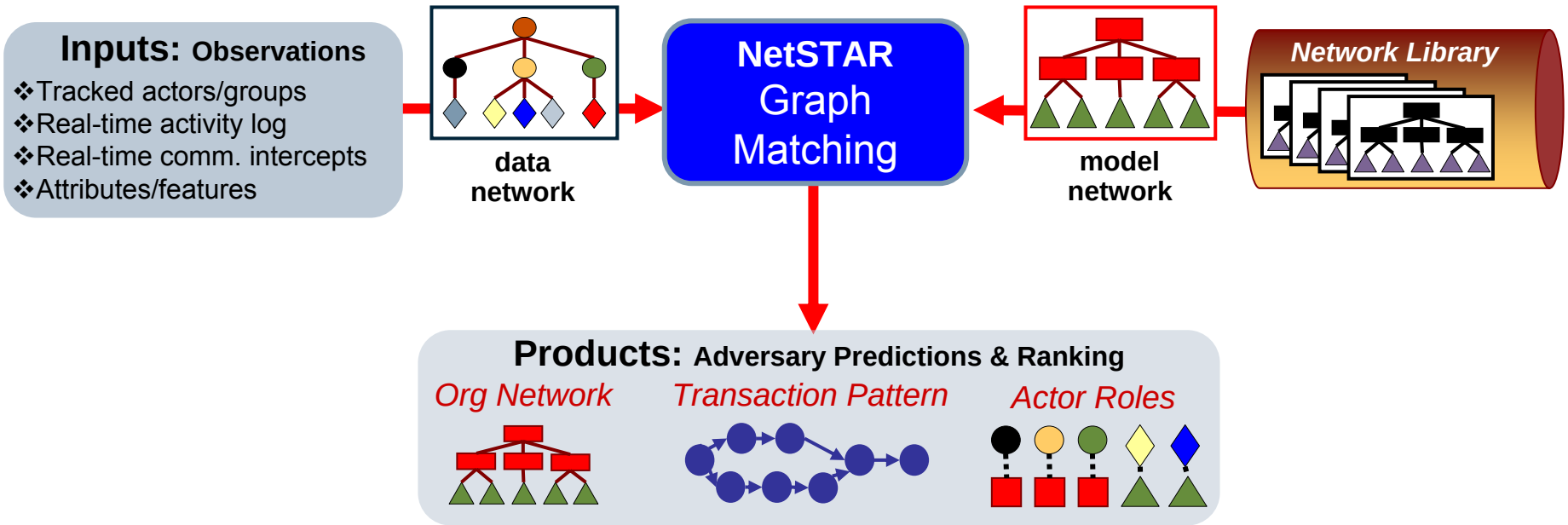


The Challenge: Uncertainty observing interactions

- **False negatives (Missing data):** unobserved transactions (modeled with miss probability)
- **False positives (Noisy data):** wrongly observed transactions or irrelevant transactions (modeled with false alarm probability)

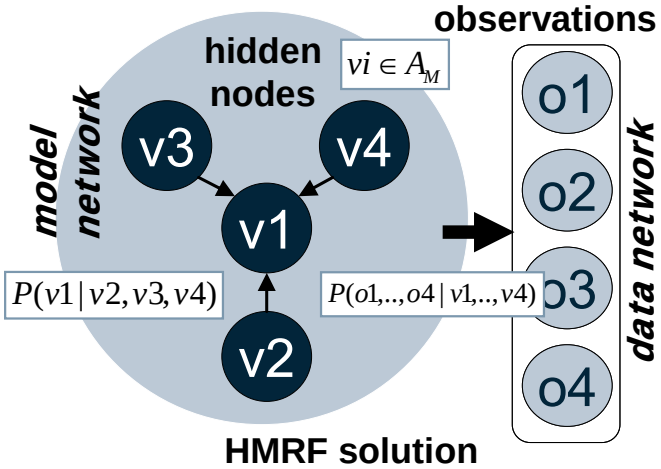


NetSTAR Solution



Problem difficulty: For 50-node network, probability of correctly identifying ≥10 (20%) nodes by chance is **1:1,000,000**

NetSTAR Model: Hidden Random Fields



Solution

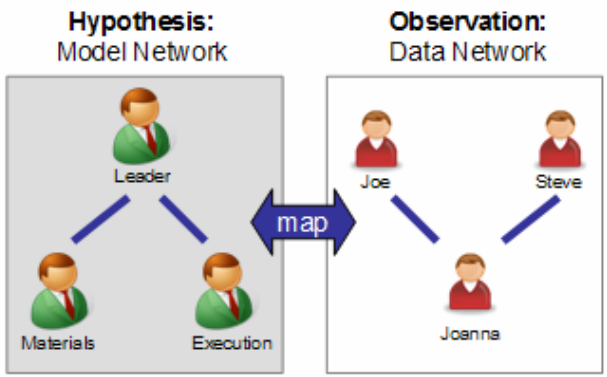
- Mapping to maximize posterior

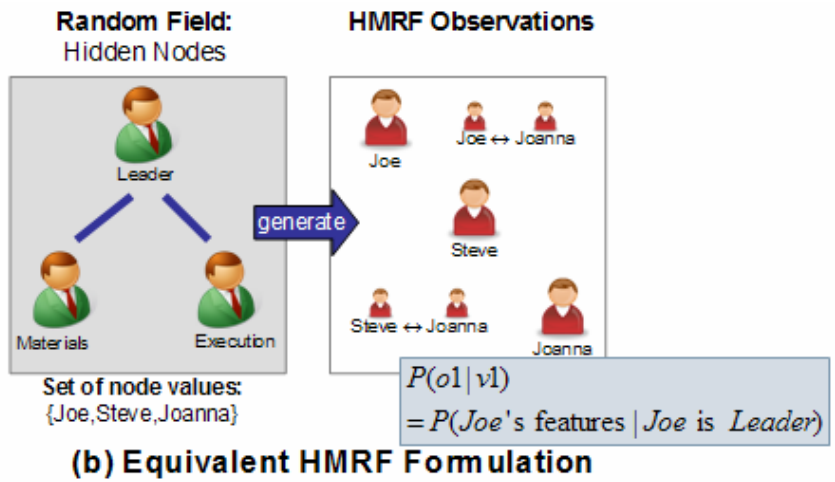
$$f^* = \arg\max_f P(f \mid G_D, G_M)$$

- Approximate posterior via energy functions due to HMRF theory

$$P(f \mid G_D, G_M) \approx \frac{1}{Z} \exp(-U(f) - U(G_D, G_M \mid f))$$

- Solve using simulated annealing
- Satisfy structural and attribute consistency





NetSTAR Advantages over Traditional Threat ID Approaches

- **Individual actor mapping**

Observed actor:

map

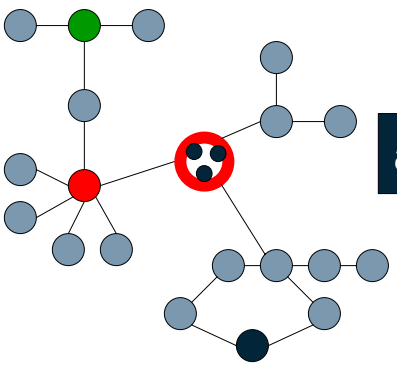
Role:
Weapons Manufacturer

Observed actions:
- Purchase fertilizer
- Cash withdrawal
- Travel to 3-rd world country

Cons: Cannot correlate coordinated actions of different actors

- **Network analysis**

Measures:
- Betweenness
- Centrality closeness/degree
- Density
- Reachability; etc



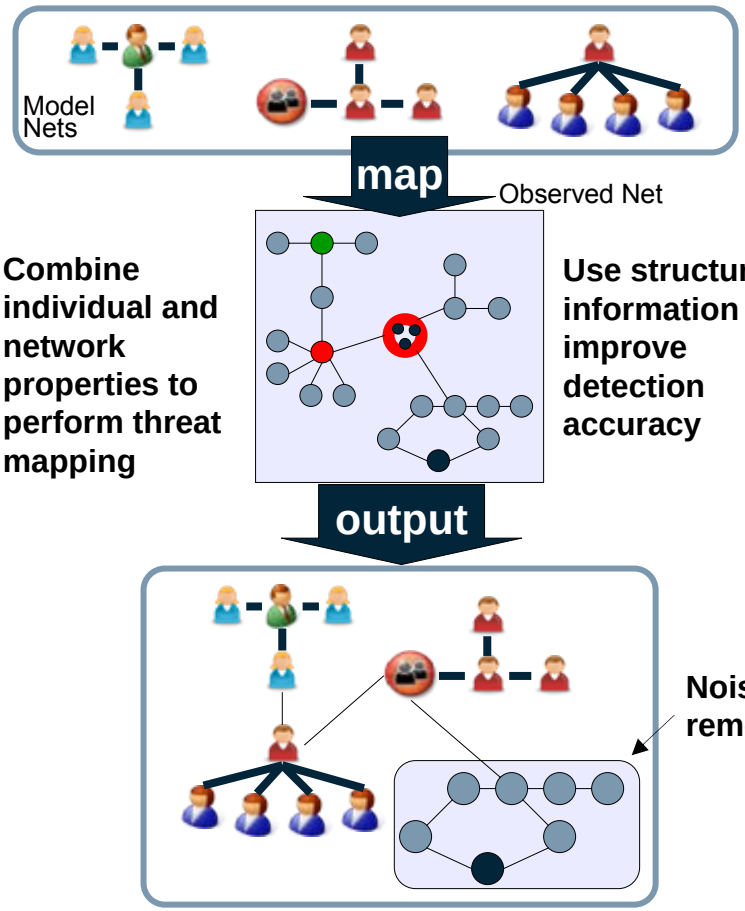
assess

-regional point person
-regional leader
-group
-liaison

Cons: Do not account for uncertainty, cannot use structural consistency, patterns embedded in metrics

- **NetSTAR**

Use historic experiences/ prior knowledge/hypotheses

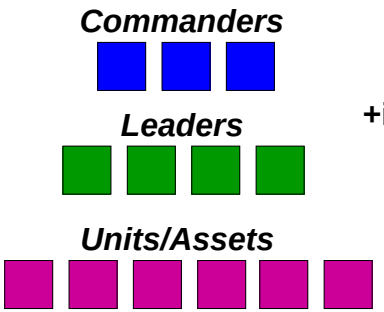


Extract "signal" from noise – missing data (false negatives), errors, deceptions (false positives)

APTIMA HUMAN-CENTERED ENGINEERING

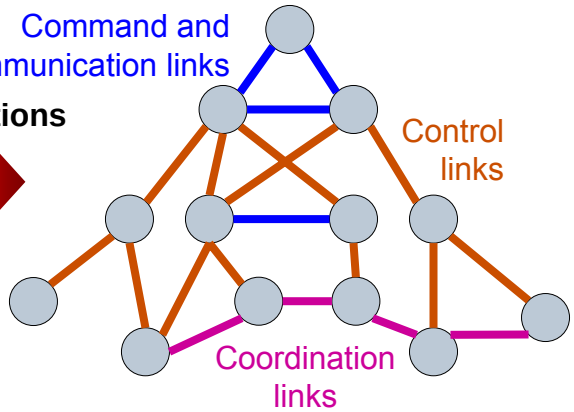
Experiment Test Networks:
Key leaders and network interactions

Ground Truth Entities



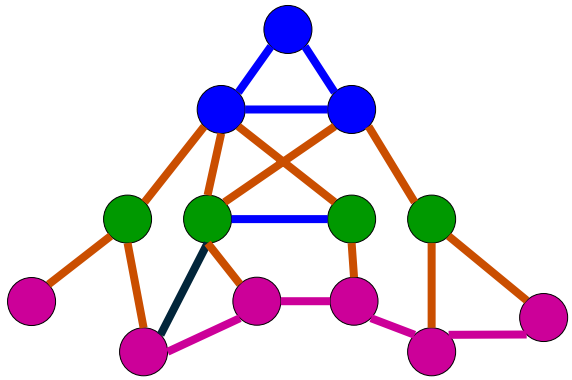
+interactions

Observed Interactions



+actions

Labeled ("colored") Network

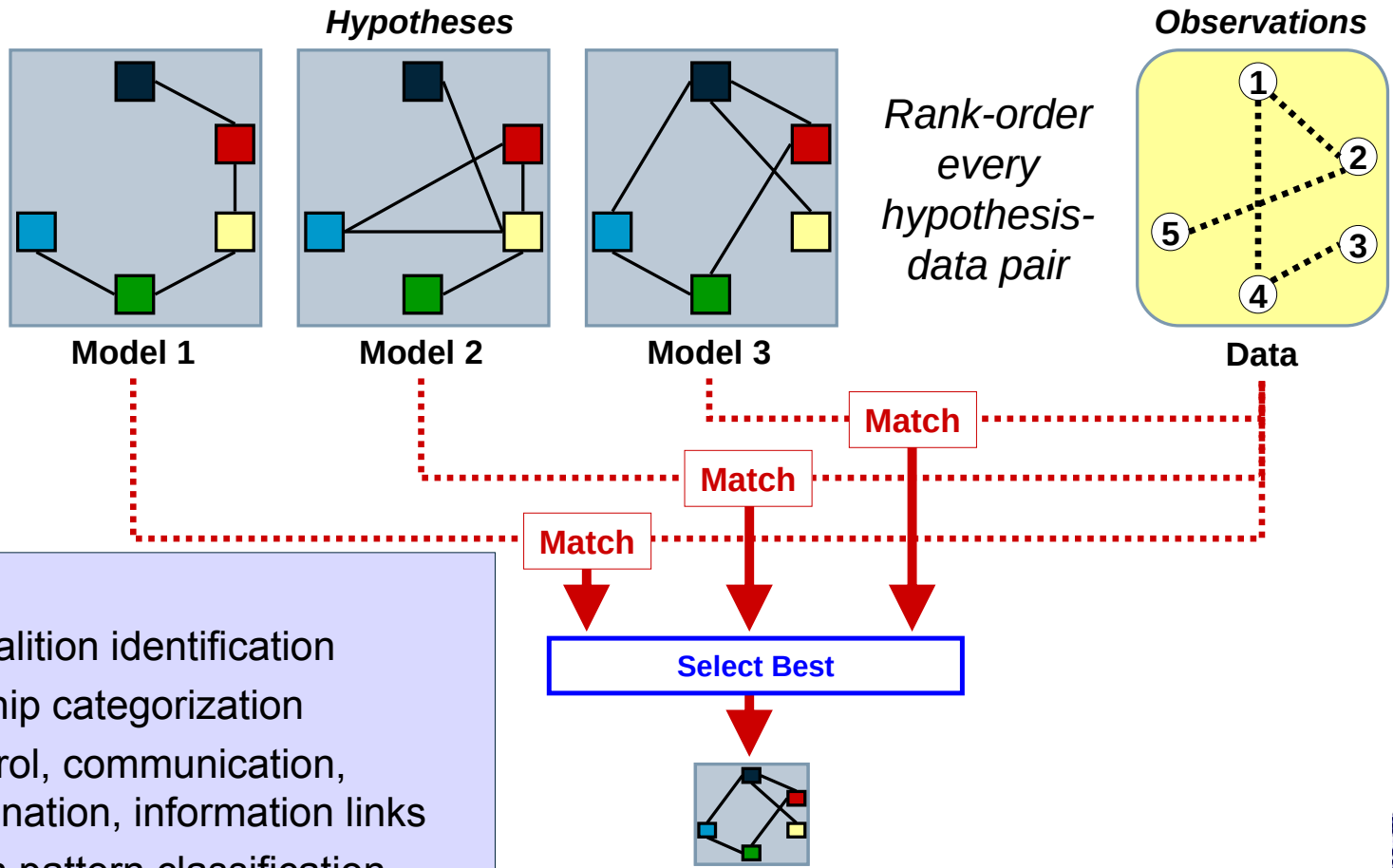


Labels: Vectors of values on links & nodes for quantitatively weighing multiple relationships and transaction types

Object	Meaning	Attributes	Observations (real world equivalent)
Communication Link	Who talks to whom about what	Classes of messages	Message between actors and message class/category (e.g., from text classification)
Control Link	Who controls/ commands whom	Types of commands issued	Commands sent from CMDR to asset; from leader to asset
Coordination Link	Who works with whom	Classes of tasks or engagements	Joint actions by multiple assets/units
Nodes	Cmdrs, Leaders, & Assets	Geographic areas of responsibility; actions performed	Task execution by actor or asset (attacks, recon)

NetSTAR Product 1: True Transaction Network

- Decide which hypothesized /model organization is active
  - From the list of alternative model org networks



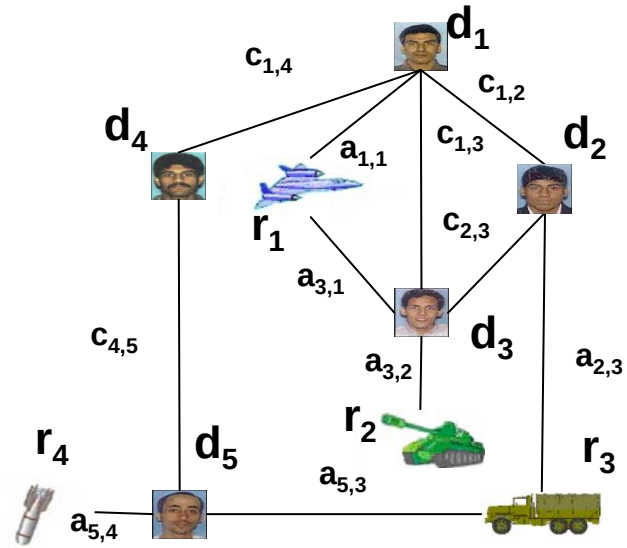
Meaning

- Group/Coalition identification
- Relationship categorization
  - Control, communication, coordination, information links
- Interaction pattern classification

APTIMA HUMAN-CENTERED ENGINEERING

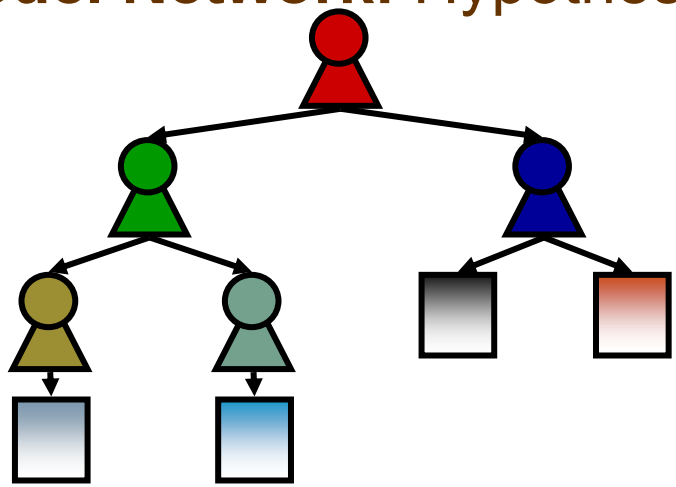
NetSTAR Product 2: Roles of Actors via Node Mapping

Data Network: Observations



- Max Structural Consistency
- Max Observed Links
- Min False Observations and Miss

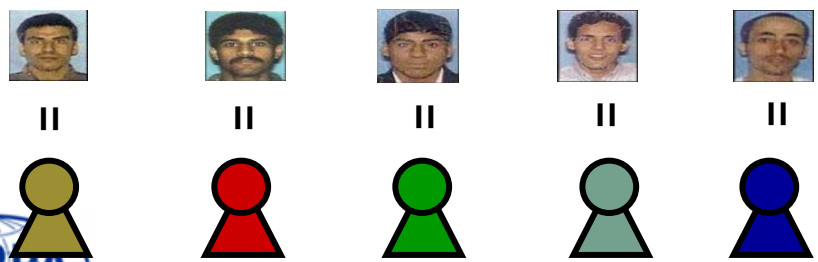
Model Network: Hypothesis



Node Mapping

Meaning

- Roles of tracked actors
- Actors' positions in the enemy org
- Actors' relationships to others



NetSTAR Experiment data flow example

observed message counts

From Name	Total Of Msg Class	CMD1	CMD2	CMD3	CMD4	CMD5	CMD6
CMD1	46		21	21	3	1	
CMD2	47	20		14	3	5	5
CMD3	36	22	7		2		5
CMD4	22	5	4	1		12	
CMD5	24	1	6		7		10
CMD6	31		5	10	3	13	

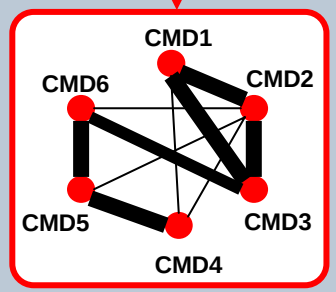
← **Raw Messages**

rule-based preprocessing

given to analysts

From Name	CMD1	CMD2	CMD3	CMD4	CMD5	CMD6
CMD1		xxx	xxx	x		
CMD2	xxx		xxx	x	x	x
CMD3	xxx	xxx				x
CMD4	x	x			xxx	
CMD5			x		xxx		xxx
CMD6		x	xxx		xxx	

network visualization

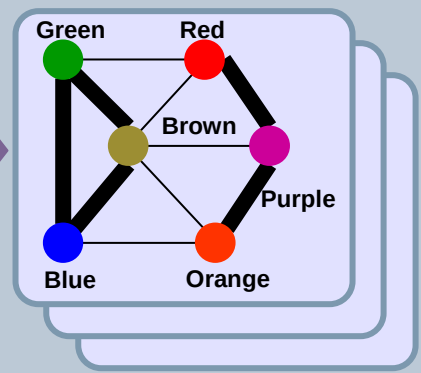


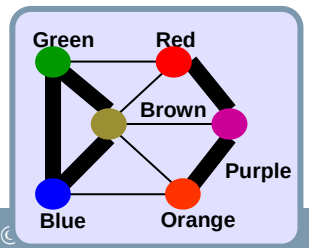
observations

map

From Name	Green	Red	Brown	Purple	Blue	Orange
Green		x	xxx		xxx	
Red	x		x	xxx		
Brown	xxx	x		x	xxx	x
Purple		xxx	x			xxx
Blue	xxx		xxx			x
Orange			x	xxx	x	

network library



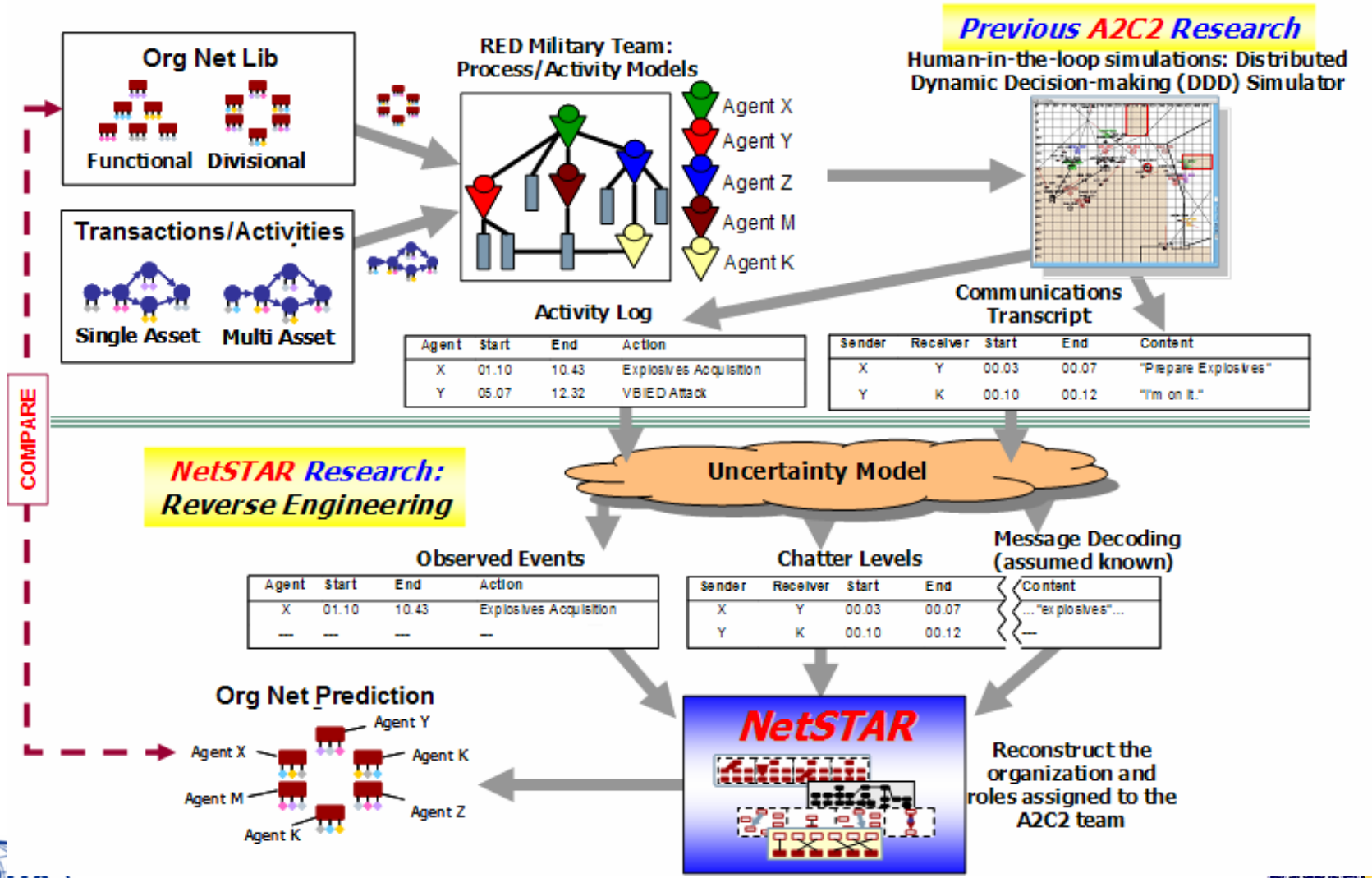


Model nodes/roles

Tracked Actors

	Green	Red	Brown	Purple	Blue	Orange
CMD1	X					
CMD2			X			
CMD3					X	
CMD4		X				
CMD5				X		
CMD6						X

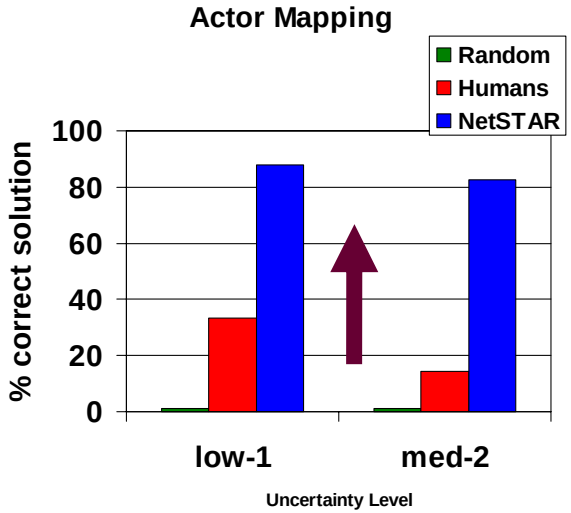
NetSTAR Validation

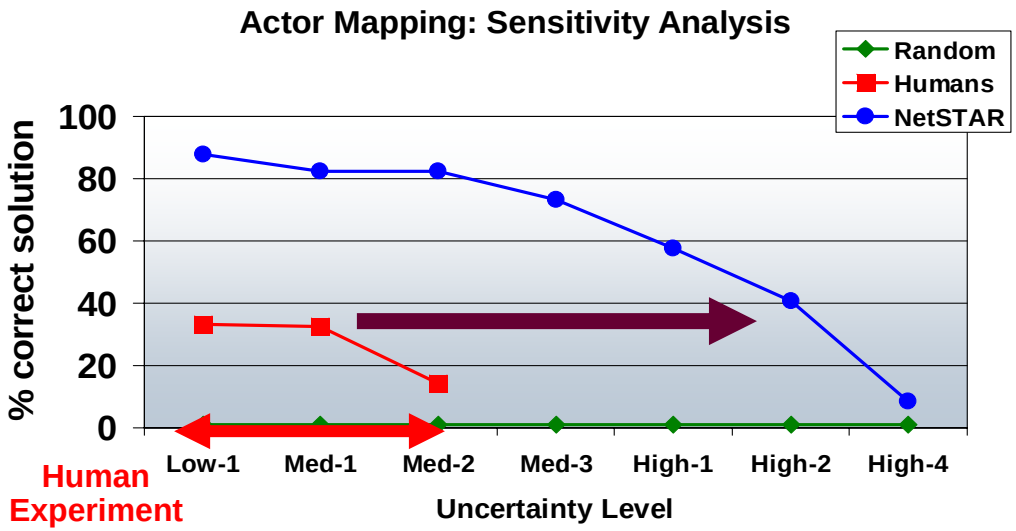


APTIMA HUMAN-CENTERED ENGINEERING

Project Findings-1:
NetSTAR Can Handle High Noise

Conducted Human Table-top Exercise and NetSTAR Algorithm Sensitivity Analyses





Uncertainty Level	Low-1	Med-1	Med-2	Med-3	High-1	High-2	High-4
% missing data	10	30	40	50	55	60	70
% deceptions/errors	10	20	30	30	30	35	45
SNR = true messages/deceptive messages	9	3.5	2	1.6666667	1.5	1.1428571	0.6666667

NetSTAR provides >2.5X better detection than human analysts under same uncertainty level

NetSTAR achieves same performance as human analysts under 3X uncertainty level

Innovation:

- C2 organizations can be distinguished by structural interaction patterns
- Algorithm solves the problem faster and more accurately than humans

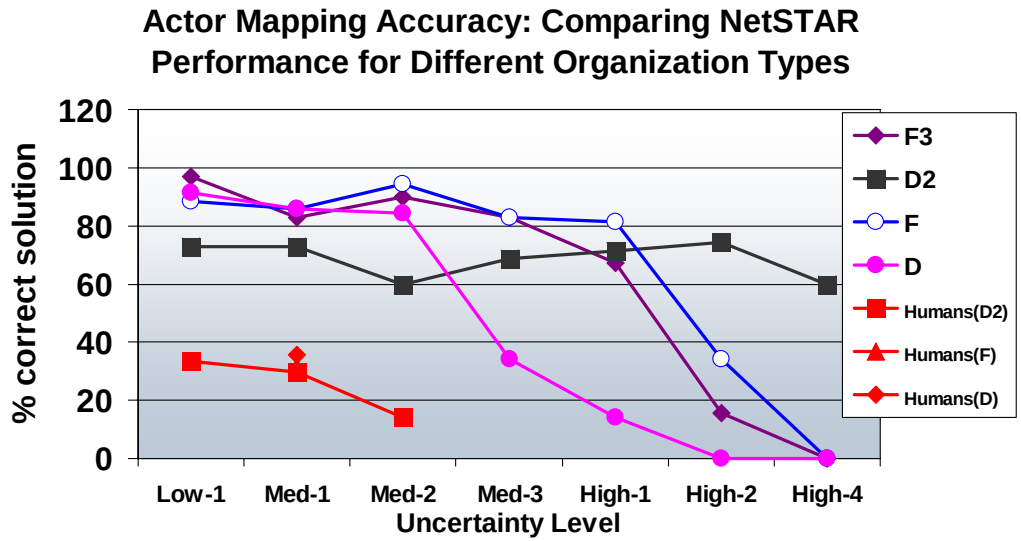
Conclusions:

- ***Actor node mapping:*** >70% correct under 50% missing data and 30% deceptions/errors
- ***Break point:*** performance degradation over 55% missing data and 35% deceptions

APTIMA HUMAN-CENTERED ENGINEERING

Project Findings-2:

NetSTAR Recognizes Unconventional Structures



Organizational types:

D = divisional organization

- CMDRs have similar resource mix & geographically distributed mission responsibilities

Real-world Example: US Army is organized divisionally

F = functional organization

- CMDRs have distinct resource mix & functionally distributed mission responsibilities

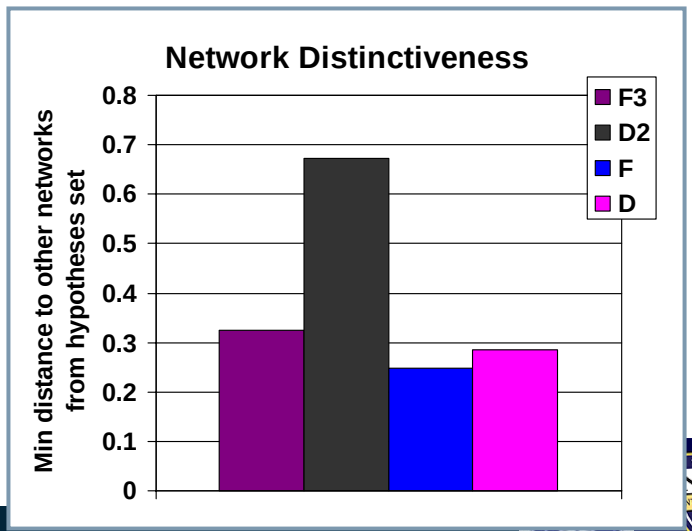
Real-world Example: US Navy is organized functionally

D2, F3 = hybrid organization

- Some CMDRs similar to D, some to F
- Current adversaries have hybrid C2 structures

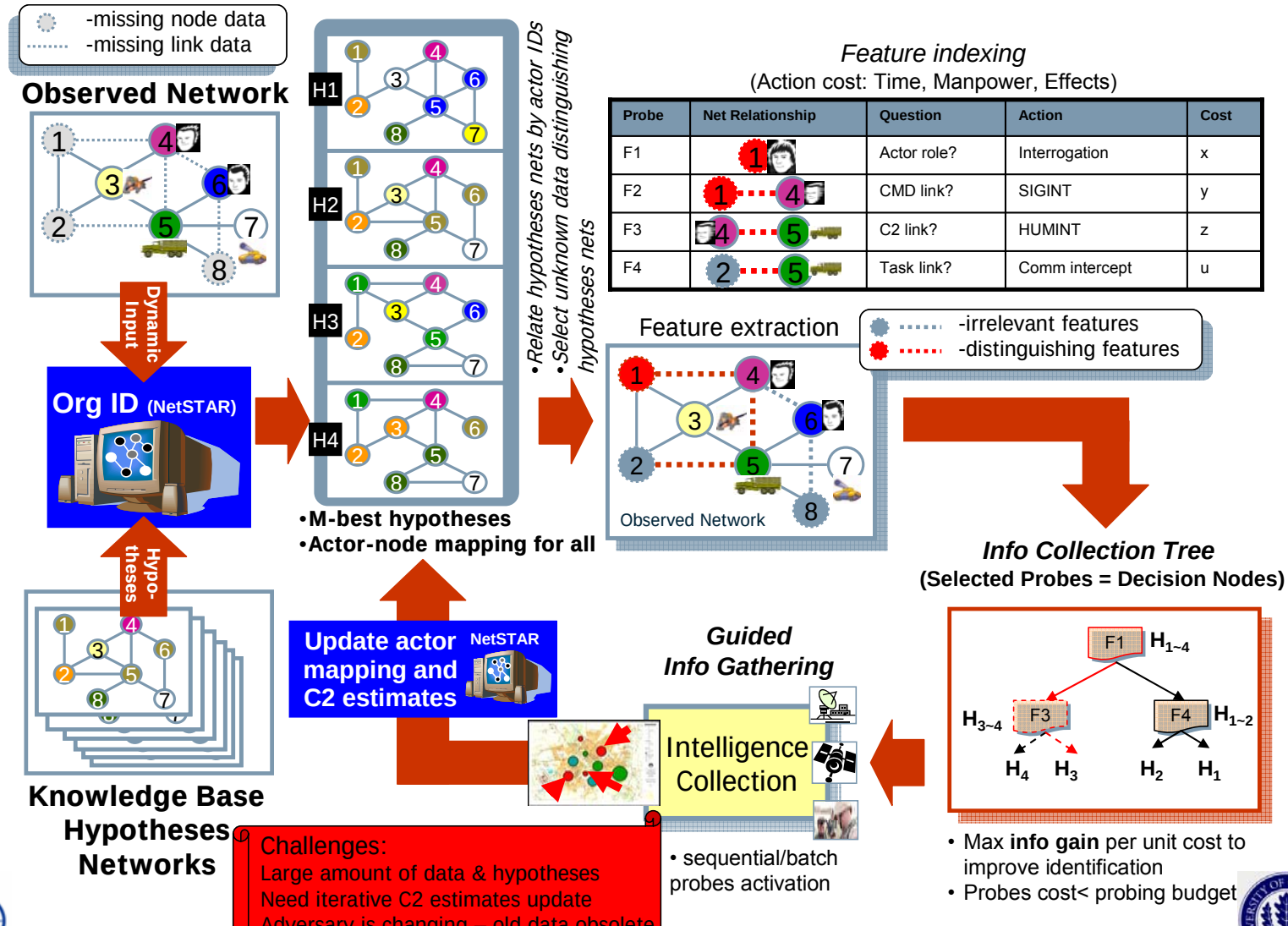
Conclusions:

- NetSTAR algorithm achieves high detection accuracy of acting **non-traditional** organizations and is not affected by experience biases
- Performance is affected by distinguishability of structures
- Some hybrid organizations exhibit unique structural patterns that enable identification



DARPA

Integrated Process:
Organization ID and Intel Planning



APTIMA HUMAN-CENTERED ENGINEERING

Details: Probes Tree Construction

- **1: Feature extraction:** Select unknown information in observed network that distinguishes current threat network hypotheses

- **2: Feature indexing:** For each feature, identify intel collection actions (probes), their cost, and ability to obtain the info (Pr of error, Pr of false alarm)

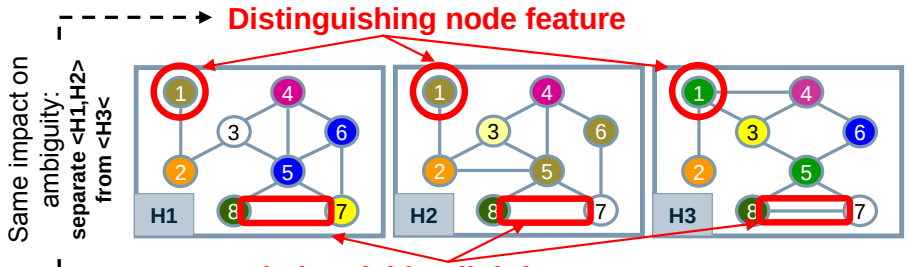
- **3: Feature organization:** Rank-order the features and organize them in a decision tree to max info gain (reduce ambiguity of current predictions) and satisfy intel collection constraints on cost of probes
  - Update probabilities for each probe's result branch

$$H(G_M \mid G_D^n, f_M) - H(G_M \mid G_D^n, f_M, O) = -\sum_{i=1}^{m} p(G_{Mi} \mid G_D^n, f_{Mi}) \log p(G_{Mi} \mid G_D^n, f_{Mi})$$

$$+ \sum_o \frac{|s : \{O = o\} \in G_{Ms}|}{m} \sum_{i=1}^{m} p(G_{Mi} \mid G_D^n, f_{Mi}, O = o) \log p(G_{Mi} \mid G_D^n, f_{Mi}, O = o)$$

$$\text{where : } p(G_{Mi} \mid G_D^n, f_{Mi}, O_k = o) = \frac{p(O_k = o \mid G_{Mi}, f_{Mi}) p(G_{Mi} \mid G_D^n, f_{Mi})}{\sum_{j=1}^{m} p(O_k = o \mid G_{Mj}, f_{Mj}) p(G_{Mj} \mid G_D^n, f_{Mi})}$$

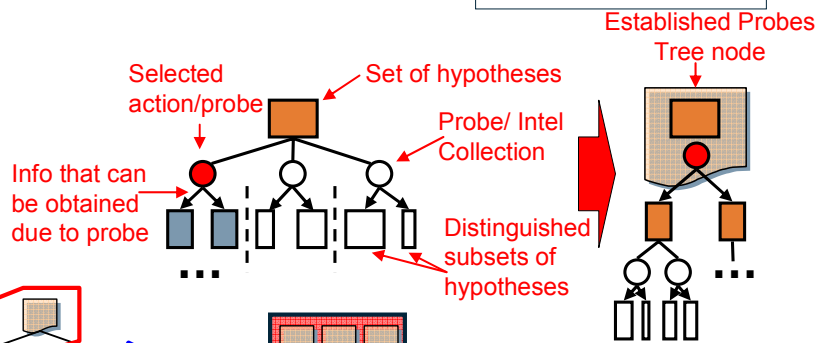
- **4: Feature clustering:** Merge related probes for integrated intelligence collection actions

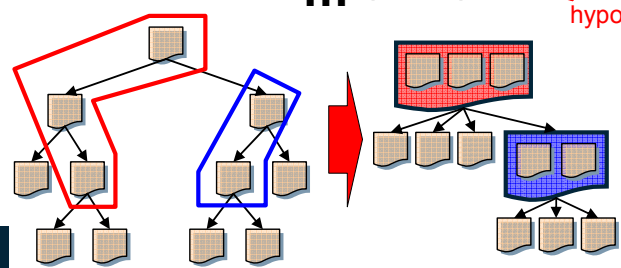


Feature	Net Relationship	Question	Action/probe	Required res	Cost
F1	1	Actor role?	Interrogation	A,B	x
F2	1 4	CMD link?	SIGINT	B	y
F3	4 5	C2 link?	HUMINT	C,E	z
F4	2 5	Task link?	Comm intercept	G,H,K	u

$$\max_k \left(H(G_M \mid G_D^n, f_M) - H(G_M \mid G_D^n, f_M, O_k) \right)$$

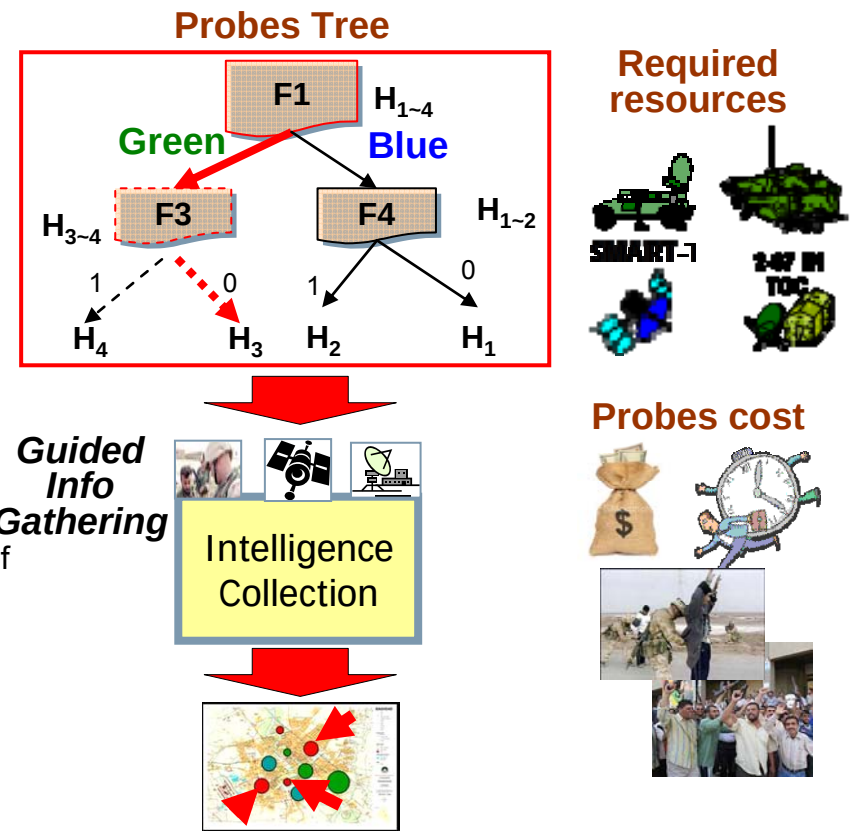
Subject to {probes cost}≤budget





Details: Intelligence Collection using Probes Tree

- **1: Resource check**
  - Is database accessible at the moment?
  - Are human collection teams available?
  - What can be consequences of intelligence collection activity?
- **2: Probes selection**
  - Select most efficient probe (e.g., intel collection to acquire F1 = interrogation to elicit role of actor 1)
- **3: Observation**
  - Obtain results from probe/intelligence gathering (e.g., role of agent 1 is Green)
- **4: Update**
  - Move to next step in probes tree
  - Update likelihoods
  - Recalculate estimated cost of intel collection plan
- **5: Repeat**
  - Next probe = feature F3 (establish existence of resource control between 4 and 5 from HUMINT)
  - Observe = F3=0 (no resource control relationship)
  - Outcome = correct adversarial network is H3



Powered by TEAMS-RDS®

Details: Updating Network Predictions

- **1: Org ID**
  - Have mission observations
  - Obtain best hypothetical/predicted networks of the enemy
  - Rank-order enemy C2 networks and obtain network actor-node mapping

$$\text{a-posteriori}: p(f_M \mid G_M, G_D^n)$$
$$\text{likelihood}: p(G_M \mid G_D^n, f_M)$$

- **2: Intelligence collection**
  - Obtain new observation "O"

- **3: Update probabilities**

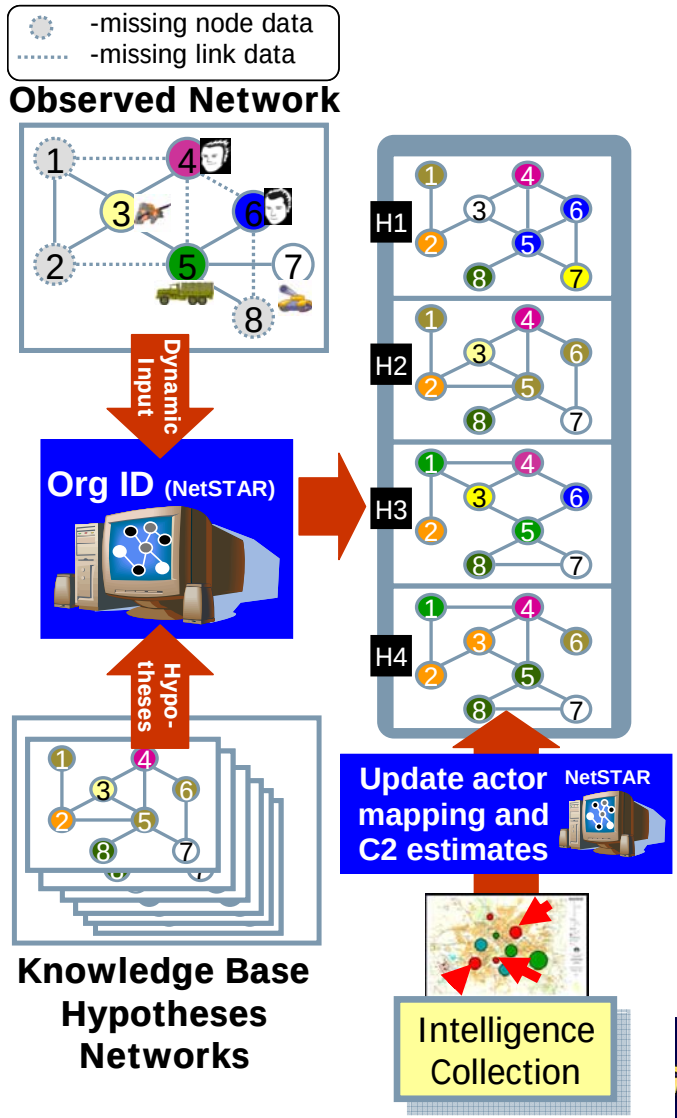
$$\text{a-posteriori}: p(f \mid G_M, G_D^n, O) \cong p(O \mid f, G_M)\, p(f \mid G_M, G_D^n)$$
$$\text{likelihood}: p(G_M \mid G_D^n, O, f) \cong p(O \mid f, G_M)\, p(G_M \mid G_D^n, f)$$

- **4: Update actor-node mapping**
  - Update energy function component

$$U(G_D^n, O, G_M \mid f) = U(G_D^n, G_M \mid f) + \log p(O \mid f, G_M)$$

  - Continue with current mapping to iteratively update best map

- **5: Update best hypotheses**
  - Check likelihood ratio for current best C2 network hypothesis



Project Conclusions

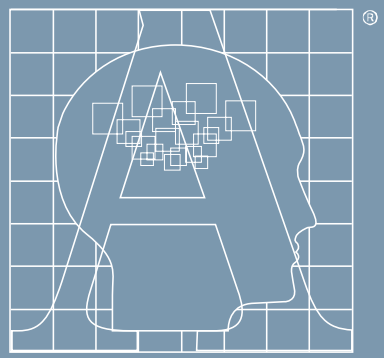
Automation

- Proven experimentally that it is possible to build automated tools that can classify network interaction patterns and identify roles of actors

NetSTAR benefits:

- Speed-up & improved accuracy of threat analysis decisions
- Handling larger volumes of data under higher uncertainty
- Increased efficiency of counteractions

Preliminary analyses indicate that the value-added of NetSTAR will be even greater for unconventional adversarial structures, such as those encountered in asymmetric warfare



APTIMA®

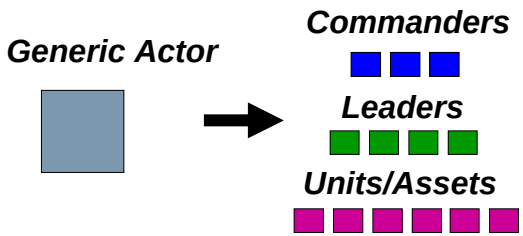
HUMAN-CENTERED
ENGINEERING

- Backups

NetSTAR Inputs:
Network Transactions Data

In NetSTAR Experiment

- **Actors**
  - Sources and targets of transactions



- **Classes of interactions**
  - Link attributes
- **Types of node roles**
  - Node attributes
- **Interaction summary**
  - # of intercepted interactions per each class per each source-target
- **Role summary**
  - # of actions or features per each type per each node

Other Applications

- Individuals
- Groups, Organizations
- Phone numbers
- Computer/email address; etc.

NetSTAR Inputs:
Network Transactions Data

In NetSTAR Experiment | Other Applications

- **Actors**
  - Sources and targets of transactions
- **Classes of interactions**
  - Link attributes
- **Types of node roles**
  - Node attributes
- **Interaction summary**
  - # of intercepted interactions per each class per each source-target
- **Role summary**
  - # of actions or features per each type per each node

Events

- Voice: Info exchange, info request, order
- Actions: Launch, Attack, Detect

→

Msg Classes

- Command
- Control
- Coordination

- Any message characteristic/classes/categories
  - Can find using text/voice classification
  - Can use duration or means of msg; etc.

NetSTAR Inputs:
Network Transactions Data

In NetSTAR Experiment | Other Applications

- **Actors**
  - Sources and targets of transactions
- **Classes of interactions**
  - Link attributes
- **Types of node roles**
  - Node attributes

Events
- Attack
- Detect
- Move

→

Roles Classes
- Task class
- Geography region

- Info about transaction source/target
  - Geolocation
  - Subnet ID
  - Size/type of group
  - Actions of target/source

- **Interaction summary**
  - # of intercepted interactions per each class per each source-target
- **Role summary**
  - # of actions or features per each type per each node

NetSTAR Inputs:
Network Transactions Data

In NetSTAR Experiment | Other Applications

- **Actors**
  - Sources and targets of transactions
- **Classes of interactions**
  - Link attributes
- **Types of node roles**
  - Node attributes

SIGINT: 20 messages between CMD1 and CMD2

- **Interaction summary**
  - # of intercepted interactions per each class per each source-target
- **Role summary**
  - # of actions or features per each type per each node

Ex: Coordination Messages Summary

Node to / Node from

From Name	Total Of Msg Class	CMD1	CMD2	CMD3	CMD4	CMD5	CMD6
CMD1	46		21	21	3	1	
CMD2	47	20		14	3	5	5
CMD3	36	22	7		2		5
CMD4	22	5	4	1		12	
CMD5	24	1	6		7		10
CMD6	31		5	10	3	13	

- Same, or qualitative summary (low/med/high)

From Name	Green	Red	Brown	Purple	Blue	Orange
Green		med	high		high	
Red	med		low	high		
Brown	high	low		low	high	med
Purple		high	low			high
Blue	high		high			low
Orange			med	high	low	

NetSTAR Inputs:
Network Transactions Data

In NetSTAR Experiment | Other Applications

- **Actors**
  - Sources and targets of transactions
- **Classes of interactions**
  - Link attributes
- **Types of node roles**
  - Node attributes
- **Interaction summary**
  - # of intercepted interactions per each class per each source-target
- **Role summary**
  - # of actions or features per each type per each node

IMINT: CMD1 detected 10 times in Village

Ex: Geo-responsibility Summary

Geographic area

Name	North Gate	Village	Market	Highway
CMD1		10	1	
CMD2	2		5	1
CMD3	21			
CMD4		2	12	
CMD5	4	2	1	2
CMD6			3	11

- Same, or qualitative summary (low/med/high)

Name	Apt Building	Public Library	Business	Secure Con
id176			high	
id221	low	low		
id309		med		
id422		low	med	
id573	high			med
id667			high	high

NetSTAR Outputs:
Interaction Pattern Classification and Actor Roles

- **Rank-order model network patterns**
  - From lib of hypothesized patterns

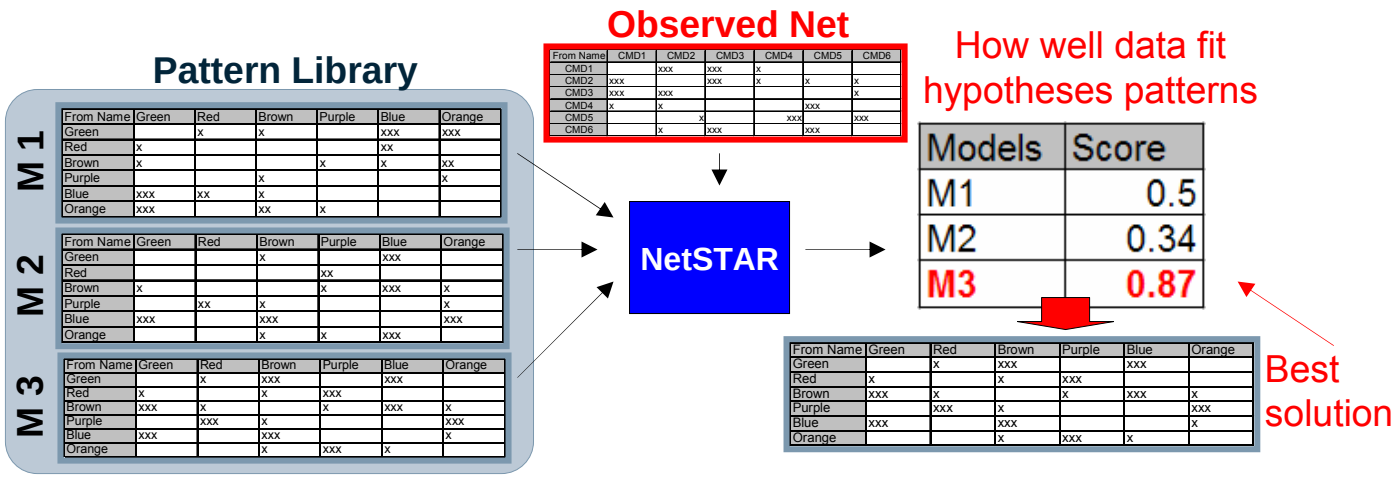
- Find "who is who"
  - Map actors to roles

In NetSTAR Experiment

- Relationship categorization
  - Control, communication, coordination, information links
- Interaction pattern classification

Other Applications

- Group/Coalition identification
- Rank any interaction pattern hypotheses



NetSTAR Outputs:
Interaction Pattern Classification and Actor Roles

In NetSTAR Experiment | Other Applications

- **Rank-order model network patterns**
  - From lib of hypothesized patterns

- **Find "who is who"**
  - Map actors to roles

- Roles & responsibilities of tracked actors
- Actors' positions in the org
- Actors' relationships to others

- "Who is who"

Observed Net

From Name	CMD1	CMD2	CMD3	CMD4	CMD5	CMD6
CMD1		xxx	xxx	x		
CMD2	xxx		xxx	x	x	x
CMD3	xxx	xxx				x
CMD4	x	x			xxx	
CMD5			x		xxx	xxx
CMD6		x	xxx		xxx	

Network Pattern

From Name	Green	Red	Brown	Purple	Blue	Orange
Green		x	xxx		xxx	
Red	x		x	xxx		
Brown	xxx	x		x	xxx	x
Purple		xxx	x			xxx
Blue	xxx		xxx			x
Orange			x	xxx	x	

NetSTAR

Node Mapping

Model nodes/roles

Tracked Actors	Green	Red	Brown	Purple	Blue	Orange
CMD1	X					
CMD2			X			
CMD3					X	
CMD4		X				
CMD5				X		
CMD6						X