Test and Evaluation of Autonomous Multi-Robot Systems

Joseph Andrew Giampapa

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA

Wednesday, 30 October 2013

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM-0000723

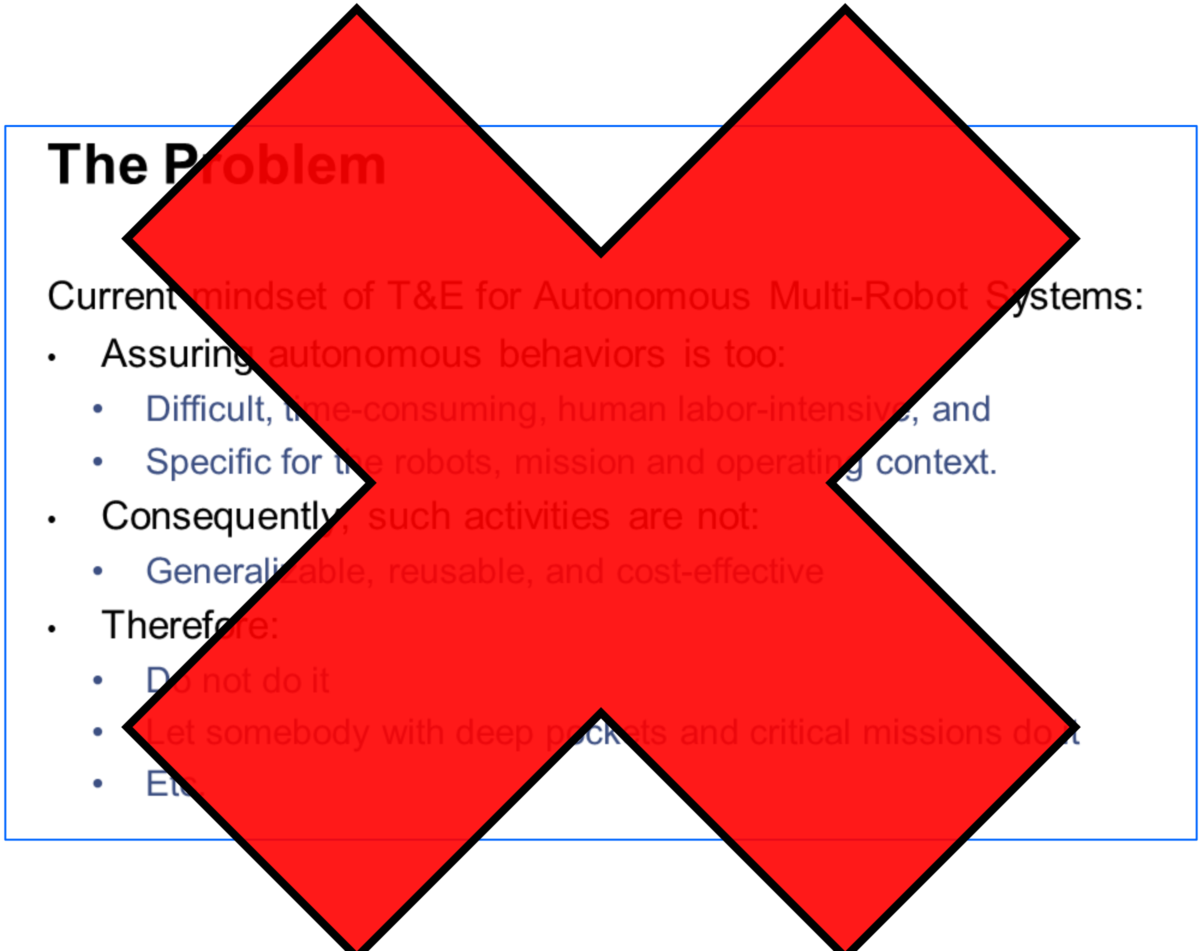
The Problem

Current mindset of T&E for Autonomous Multi-Robot Systems:

- Assuring autonomous behaviors is too:
  - Difficult, time-consuming, human labor-intensive, and
  - Specific for the robots, mission and operating context.
- Consequently, such activities are not:
  - Generalizable, reusable, and cost-effective
- Therefore:
  - Do not do it
  - Let somebody with deep pockets and critical missions do it
  - Etc.

No!

This is the wrong way to think about the problem.



Research Claim

It *is* possible to assure the future performance of an autonomous multi-robot team,

- To acceptable degrees (*justified confidence*) of expectation
- Through quantitative assurance techniques
  - *Not only is a behavior possible, but its likelihood can be estimated.*
  - Entails qualitative claims

This presentation introduces two complementary research approaches to quantitatively assuring the behaviors of autonomous multi-robot systems:

1. Probabilistic Model Checking
2. Behavioral Reliability Analysis

Outline for the Rest of the Presentation

- **Robots, Agents, Humans, …**
- Other Key Terms & Ways of Thinking
- Technical / Research Approach
- Probabilistic Model Checking
- Behavioral Reliability Analysis
- Conclusions and Take-Away Message

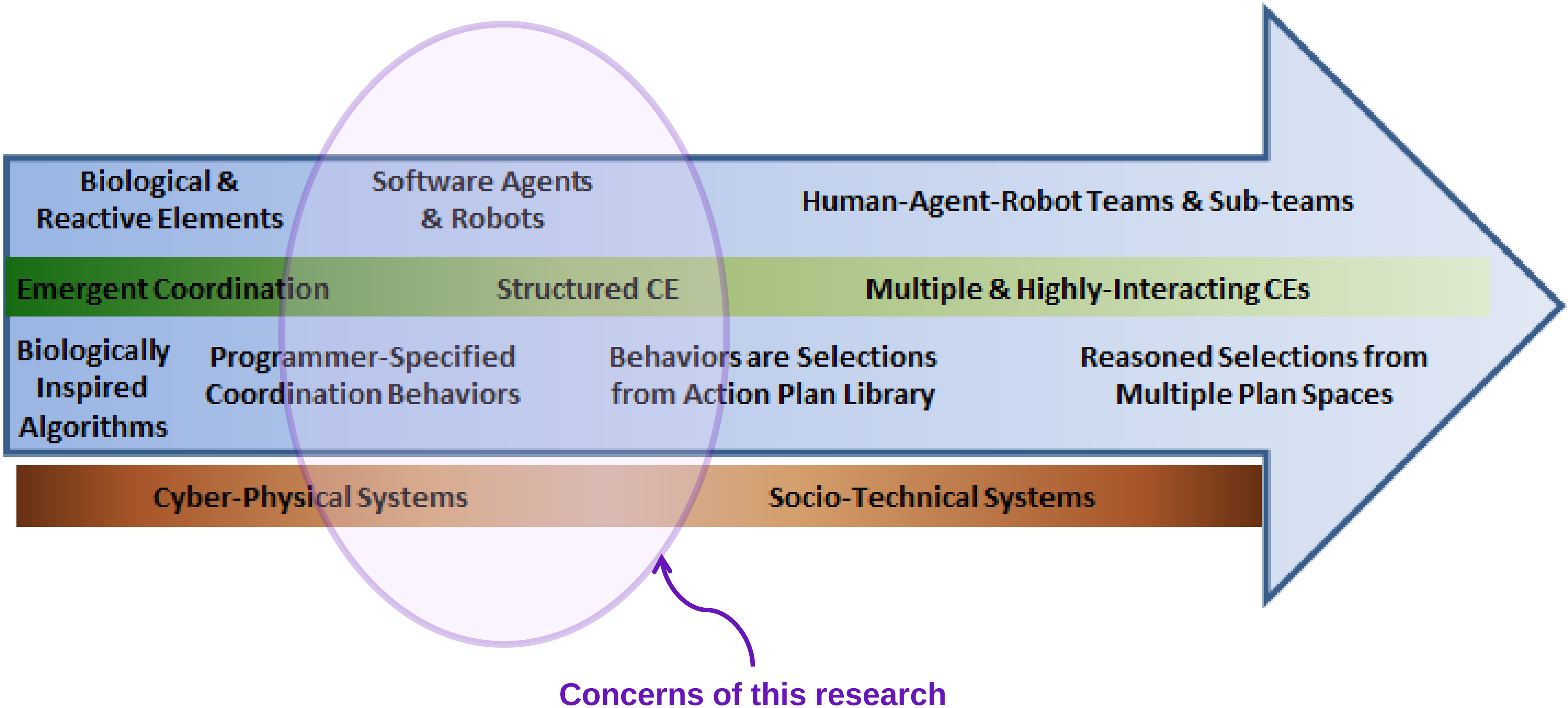
Robots, Agents, Humans, …

- Robot: autonomous, non-teleoperated cyber-physical system entity
- Agent: autonomous, socially-aware, software agent
- Autonomous system: collection of autonomous, socially-aware entities
  - Includes humans, often in a limited social role/context

- Cyber-physical system (CPS): a software system that must support and accommodate a non-trivial interaction model of the physical world
- Socio-technical system: a socially-aware, system of autonomous entities and systems

- This presentation is focused on assuring robots,
  - Treating them primarily as CPS entities
  - Assuming them to operate as members of a socio-technical system

Range of Autonomous Coordination Behaviors



CE: Coordination Element

Structured Coordination Elements

- Mission Objective
  - Can be single or multiple, of equal or subordinate rank
  - Provides:
    - The motivation for quantitative assurance claims
    - Metrics by which the claims are assessed
- Operating Context
  - All possible influences on the mission outcome
  - Examples: physical, computing, data communication environments
- Individual Capabilities
  - Union of individual agent capabilities across all members of the team
  - Quantitative evaluation of individual capabilities contribute to overall evaluation of the team.
- Team Plan
  - Specifies coordination behaviors
  - Should be designed to:
    - Ensure that team scalability can be achieved
    - Remediate deficiencies of individuals at achieving team mission objective
  - Includes: roles, sub-plan assignments of individuals and subgroups
  - Individual capabilities measured with respect to role & sub-plan assignments

Team Maintenance Coordination Elements

- Can apply to all phases of an autonomous mission, to varying degrees
- Monitor Team Performance
  - Determine if an objective will not be met or if approach can be improved
- Detect Failure
  - The team must agree that there is a failure that needs to be addressed
  - Otherwise, there can be partial failures and cascading de-commitments
- (Optional) Repair
  - Recruit additional team members
    - Either in substitution, or
    - To enrich the performance of the team
  - Adopt a new team plan
  - Reassign roles and transition to them
- Consensual Team Plan Termination
  - Can be due to a detected team failure, or
  - Due to completion of the team plan

Outline for the Rest of the Presentation

- Robots, Agents, Humans, …
- **Other Key Terms & Ways of Thinking**
- Technical / Research Approach
- Probabilistic Model Checking
- Behavioral Reliability Analysis
- Conclusions and Take-Away Message

We Desire Robots & Autonomous Systems

1. As an extension to the human individual; to permit humans to:
   - *Perceive* and *do* more
   - Have a more comprehensive awareness of:
     - The environment that needs to be accurately perceived
     - Other operations of (potential) relevance to their mission
2. As a projection of the human individual in dangerous environments
   - IED investigations
   - Reconnaissance of hostile terrain
   - Checkpoint manning

Autonomy

- The property of an entity to have persistent, goal-directed behavior
  - **Goal-directed behavior** allows for alternative courses of action to achieving a goal in a dynamic and unpredictable environment
    - Environmental dynamism and unpredictability render autonomous systems difficult to assure.
  - **Persistence** ensures the entity will attempt to achieve the goal as long as:
    - It has an interest in attempting to do so
    - It can reason that it has the means of achieving it
      - Reasoning can involve relying on commitments from other entities to assist it in achieving its goal
      - Implies cooperative disposition, which is characterized by the range:
        - Altruism – self-interest
- Requires the ability to sense and interpret the environment in the context of an outcome space
  - Interpretation is given by software that maps sensory input to an action or consequence
- Presumes that the range of possible actions / consequences is known
  - Implies a computational plan library

Autonomous Coordination

- Ensures that a goal can be achieved collectively by a group
  - If goal is achievable
  - If individuals of a group can contribute capabilities for goal achievement
    - And they commit to using those capabilities to achieve it
  - Through plan repair and response to environmental dynamics
- Means of adapting collective plan for achieving team goal
- Achieves scalability by managing resources
  - Time and space
  - Avoiding collisions
  - Synchronizes power drives
  - Multi-tasking and parallelizing work

Autonomous Multi-Robot Coordination

- Individual robots with targeted functionalities typically perform better
  - They are more reliable and cost-effective than more general-purpose systems that attempt to do many things.
  - Address trade-off decisions by fielding multiple versions
    - Rather than one “Transformer” robot that can be both big and small
    - Deploy two robots: one big and one small
- Achieves goals that are beyond the capability of an individual
- Maximizes performance quality attributes via system scaling
  - Increase likelihood of mission success by adding more entities
  - Decrease mission time requirements by distributing tasks
- Minimizes inter-entity interference by managing shared resources
  - Collision avoidance (shared temporal-spatial state)
- Individual role assignments can remediate individual deficiencies
  - In serially-coordinated systems, entities that drift most, deploy last

Assurance

- An analysis technique by which trust in a system is established
  - State claims about properties of a system
  - Prove those claims via reasoned arguments
  - Accept proof when:
    - A knowledgeable reviewer
    - Can read assurance claims and arguments to support them
    - Have justified confidence in the expected behavior of the system
      - Justified confidence implies engineering tolerances
- Testing and evaluation (T&E)
  - A synonym for “assurance”
  - *When* “testing and evaluation” and “assurance” are *used together*:
    - Assurance applies more broadly: any form of claim or logical argument
    - T&E sometimes refers to specific assumptions & context; in checklist form
- Quantitative methods for assuring systems
  - Allow numeric measures to be made of them

Trust for Autonomous Systems

Two general comments:

1. Predictive trust must be evaluated in the context of a comprehensive assurance argument.
    - It is insufficient to just use results of test suites in arenas and field tests.
    - Assumptions and their limits need to be understood
2. Dynamic trust depends on the way in which system state and decision support rationale are presented to the human participants in the system.
    - Legal reasoning systems present the facts and logic that support the case.
    - Wealth of studies in human factors for trust in automation
        - Issues range from:
            - Over-reliance, to
            - Not understanding constraint space,
            - Among others

Outline for the Rest of the Presentation

- Robots, Agents, Humans, …
- Other Key Terms & Ways of Thinking
- **Technical / Research Approach**
- Probabilistic Model Checking
- Behavioral Reliability Analysis
- Conclusions and Take-Away Message

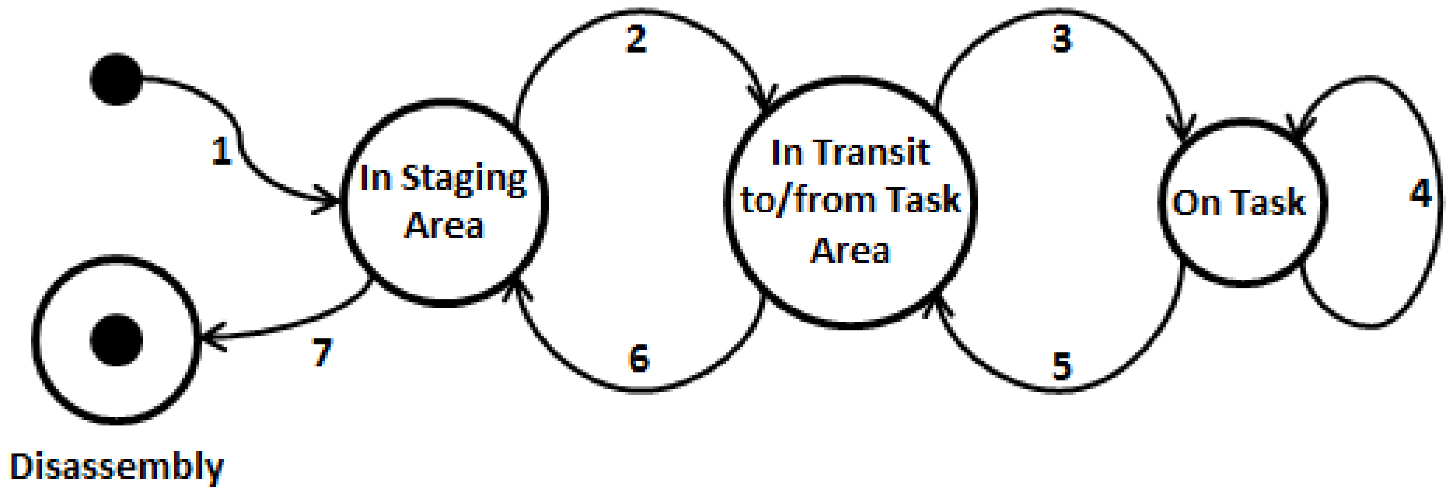
Technical / Research Approach

- Identify autonomous behaviors that contribute to mission success.
  - Follow a scan pattern, identify simulated mines in terrain, etc.
- Determine metrics by which such behaviors can be quantified.
  - Likelihood of passing through center of a terrain cell; % of terrain covered
  - Mine ID metrics: accuracy, precision, recall, F-measures, combination of evidence, etc.
  - Time to complete mission; Likelihood of recovering N robots of team
- Evaluate the behaviors:
  - Atomistically, for each individual robot
  - In an operating context
  - In coordination with other autonomous entities
- Find models that relate atomistic performance to overall coordinated performance
  - Probabilistic Model Checking
    - Discrete Time Markov Chain (DTMC)
    - alpha-Probabilistic Automaton (αPA)
    - Linear Temporal Logic (LTL)
    - Cumulative reward within the DTMC / αPA
  - Reliability Analysis
    - Express relation as a conditional probability of performance metrics given physical features
      - CPS features: mines, way points, terrain type

Rationale for Model Design Criteria

- Maximize reuse by identifying variables
  - Known Variables
    - Individual performance (identified a priori)
    - Physical and computing context (identified a priori)
    - Features that reliably predict performance (discovered)
  - Unknown Variables / Difficult to Characterize
    - Reasons for deviations from expected behavior
    - Joint cyber-physical interactions with other robots
- Compositionality requires (near-)independence
  - Design models to maximize independence
  - Identify the “principal components” of a full mission
    - These are usually general for a type of mission
  - Complete independence is not always necessary
    - A specific joint state space might have little impact on overall mission

Phases of an Autonomous Mission



Missions often evolve in the following phases:

1. Assembly in a staging area
2. Travel to task area from assembly area
3. Ingress into task area & transition to physical roles
4. Performance of task
5. Travel from task area to departure corridor
6. Return to staging area
7. Disassembly

Outline for the Rest of the Presentation

- Robots, Agents, Humans, …
- Other Key Terms & Ways of Thinking
- Technical / Research Approach
- **Probabilistic Model Checking**
- Behavioral Reliability Analysis
- Conclusions and Take-Away Message

Project Team

Assuring Distributed Autonomous Coordination (ADAC)

- Sagar Chaki
- Joseph A. Giampapa
- David S. Kyle
- Occasional external collaborators:
    - Edmund Clarke, CMU Computer Science Department
    - Arie Gurfinkel
    - Anvesh Komuravelli, CMU Computer Science Department
    - Paul Scerri, CMU Robotics Institute

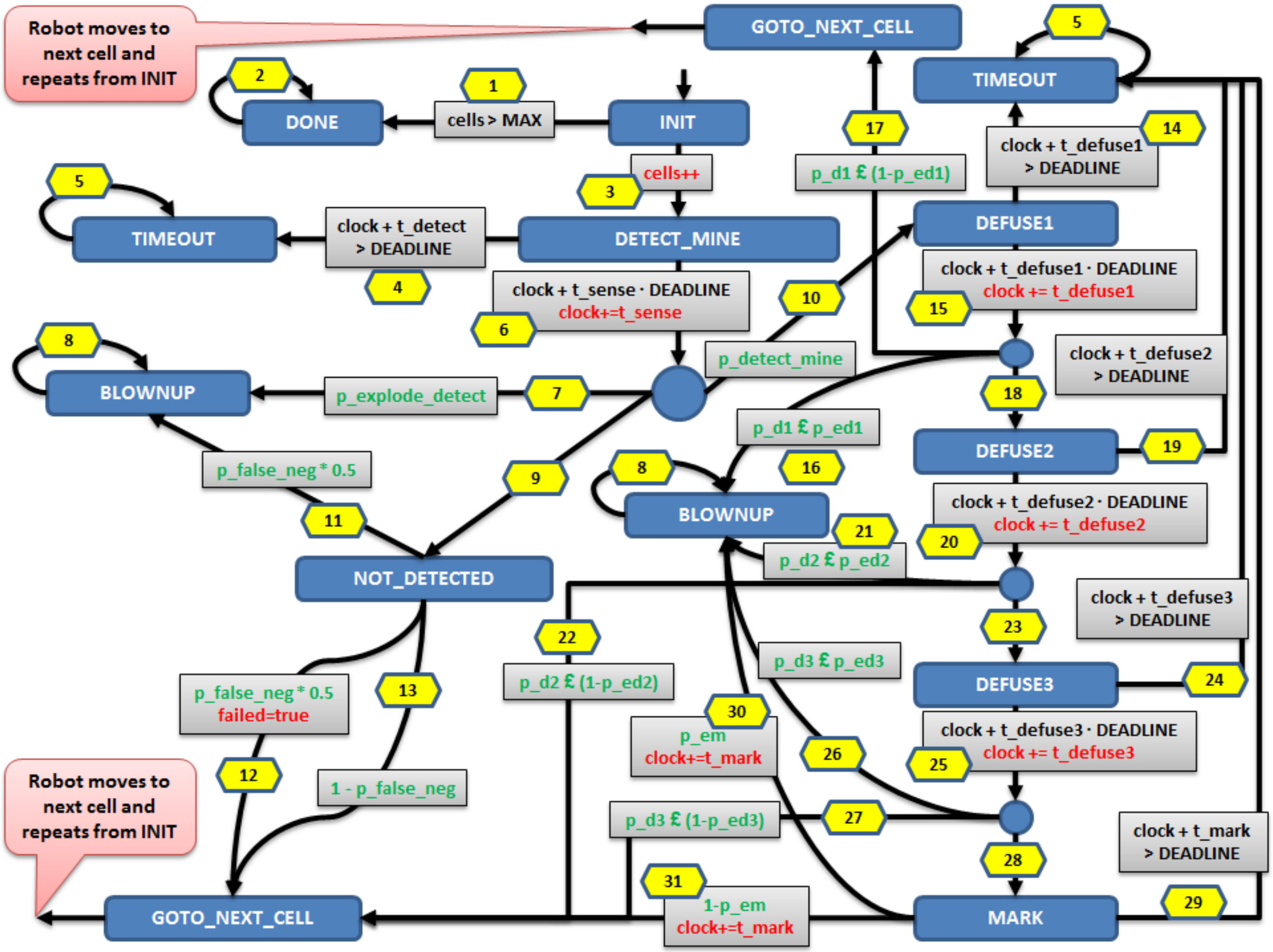
Classical Model Checking

- Classical model checking:
  - Given:
    - A system, *S*, represented by a state space
    - Property, *p*, to prove
    - Assertions of state at a given time, *t*
  - Prove:
    - A path exists from initial state to terminal state, in which
    - The property is true
- Advantage: exhaustive exploration of state space
- Disadvantages:
  - State space explosion
  - Atomistic representation of state space
- Weaknesses for robotic systems
  - Both *S* and *p* are stochastic
    - Difficult to enumerate all states & properties without abstraction
    - Risk of state space explosion
  - More useful to consider the *likelihood* of *p* being true

Probabilistic Model Checking

- Ways of expressing state *S* for stochastic systems
  - Discrete or continuous time Markov chain (DTMC, CTMC)
  - Markov decision process (MDP)
  - Probabilistic timed automaton (PTA)
- Expression of property *p* for stochastic systems
  - Probabilistic temporal logic
    - e.g. PCTL – probabilistic computation tree logic
- PRISM, the probabilistic model checker we used
  - MTBDDs – multi-terminal binary decision diagrams
  - PCTL – probabilistic computation tree logic
  - Individual robot's state modeled as DTMC (discrete time Markov chain)
- Additionally, we used:
  - Abstraction & segmentation, according to our models
  - Reliability engineering rules of combination to reduce state space

Probabilistic Model



Coordination Strategies

- C0: Parallel Independent
  - Each robot assigned cells to de-mine
  - Each cell is allocated to exactly one robot
  - Robots:
    - Work independently
    - Stop after demining the cells allocated to them
- C1: Follow the Leader (hot-standby dynamic substitution)
  - All robots move together as a team
    - One leader in front
    - Rest follow
  - Leader does all the work:
    - Detection, defusion, marking
  - If leader disabled by explosion
    - A follower is promoted to leader

Metrics of Mission Success

1. *Succ* = Probability of covering all cells
   a. Without blowing up, or
   b. Without missing a mine
2. *Cov* = Expected number of cells
   a. Defused, or
   b. Marked as containing a mine
3. Time used:
   a. Within a DTMC, as a deadline
   b. To provide initial synchronization of multiple DTMCs

Mission Success

	succ			
N	A0	A1	A2	A3
1	1.51E-05	2.62E-05	0.3659	0.6026
2	1.08E-04	1.72E-04	0.6430	0.7570
3	3.90E-04	5.73E-04	0.7468	0.7766
4	9.57E-04	1.30E-03	0.7724	0.7782
5	1.80E-03	2.29E-03	0.7771	0.7783
6	2.80E-03	3.34E-03	0.7778	0.7783
7	3.77E-03	4.27E-03	0.7779	0.7783
8	4.58E-03	4.96E-03	0.7779	0.7783
9	5.15E-03	5.41E-03	0.7779	0.7783
10	5.51E-03	5.67E-03	0.7779	0.7783

Mission Success (succ) with increasing number of robots (N).

A0: P(detect)=low, P(defuse)=low

A1: P(detect)=low, P(defuse)=high

A2: P(detect)=high, P(defuse)=low

A3: P(detect)=high, P(defuse)=high

Mission Success

1. **Detecting a mine contributes more to mission success than defusing a mine.**
2. **Maximum likelihood of success is achieved with 5 robots. Near-maximum success with 3.**

		succ		
			A2	A3
			0.3659	0.6026
			0.6430	0.7570
[illegible]	[illegible]	[illegible]	0.7468	0.7766
4	9.57E-04	[illegible]	0.7724	0.7782
5	1.80E-03	2.29E-03	0.7771	0.7783
6	2.80E-03	3.34E-03	0.7778	0.7783
7	3.77E-03	4.27E-03	0.7779	0.7783
8	4.58E-03	4.96E-03	0.7779	0.7783
9	5.15E-03	5.41E-03	0.7779	0.7783
10	5.51E-03	5.67E-03	0.7779	0.7783

Mission Success (succ) with increasing number of robots (N).

A0: P(detect)=low, P(defuse)=low

A1: P(detect)=low, P(defuse)=high

A2: P(detect)=high, P(defuse)=low

A3: P(detect)=high, P(defuse)=high

Terrain Coverage

	cov							
	C0				**C1**			
N	A0	A1	A2	A3	A0	A1	A2	A3
1	18.089	19.827	70.525	88.362	18.089	19.827	70.525	88.362
2	34.168	36.841	83.646	93.997	35.794	39.022	93.573	99.057
3	45.613	48.455	88.558	95.903	52.427	56.595	98.881	99.931
4	54.998	57.728	91.495	96.997	67.037	71.392	99.813	99.984
5	61.538	64.077	93.184	97.613	78.799	82.646	99.945	99.987
6	66.028	68.388	94.218	97.984	87.380	90.316	99.960	99.987
7	69.289	71.497	94.916	98.233	93.029	94.996	99.962	99.987
8	72.788	74.812	95.620	98.483	96.388	97.559	99.962	99.987
9	74.631	76.551	95.975	98.608	98.195	98.825	99.962	99.987
10	78.522	80.204	96.691	98.859	99.079	99.392	99.962	99.987

Terrain coverage (cov) with increasing number of robots (N).

C0: Follow leader; substitute failed robot (no coordination)

C1: Follow leader; scan assigned role.

A0: P(detect)=low, P(defuse)=low

A1: P(detect)=low, P(defuse)=high

A2: P(detect)=high, P(defuse)=low

A3: P(detect)=high, P(defuse)=high

Terrain Coverage

Only 2 robots needed for Mission Success with:

1. **Follow-the leader coordination**
2. **Individual robots have high detection capability**

					C1			
							A2	A3
							70.525	88.362
							[illegible]	99.057
							98.881	99.931
4	54.998	57.728	91.495	96.997	67.037	71.392	99.813	99.984
5	61.538	64.077	93.184	97.613	78.799	82.646	99.945	99.987
6	66.028	68.388	94.218	97.984	87.380	90.316	99.960	99.987
7	69.289	71.497	94.916	98.233	93.029	94.996	99.962	99.987
8	72.788	74.812	95.620	98.483	96.388	97.559	99.962	99.987
9	74.631	76.551	95.975	98.608	98.195	98.825	99.962	99.987
10	78.522	80.204	96.691	98.859	99.079	99.392	99.962	99.987

Terrain coverage (cov) with increasing number of robots (N).

C0: Follow leader; substitute failed robot (no coordination)

C1: Follow leader; scan assigned role.

A0: P(detect)=low, P(defuse)=low

A1: P(detect)=low, P(defuse)=high

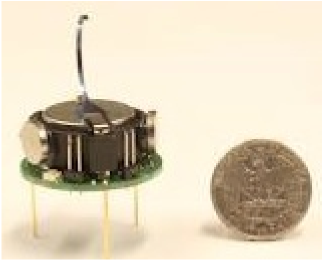
A2: P(detect)=high, P(defuse)=low

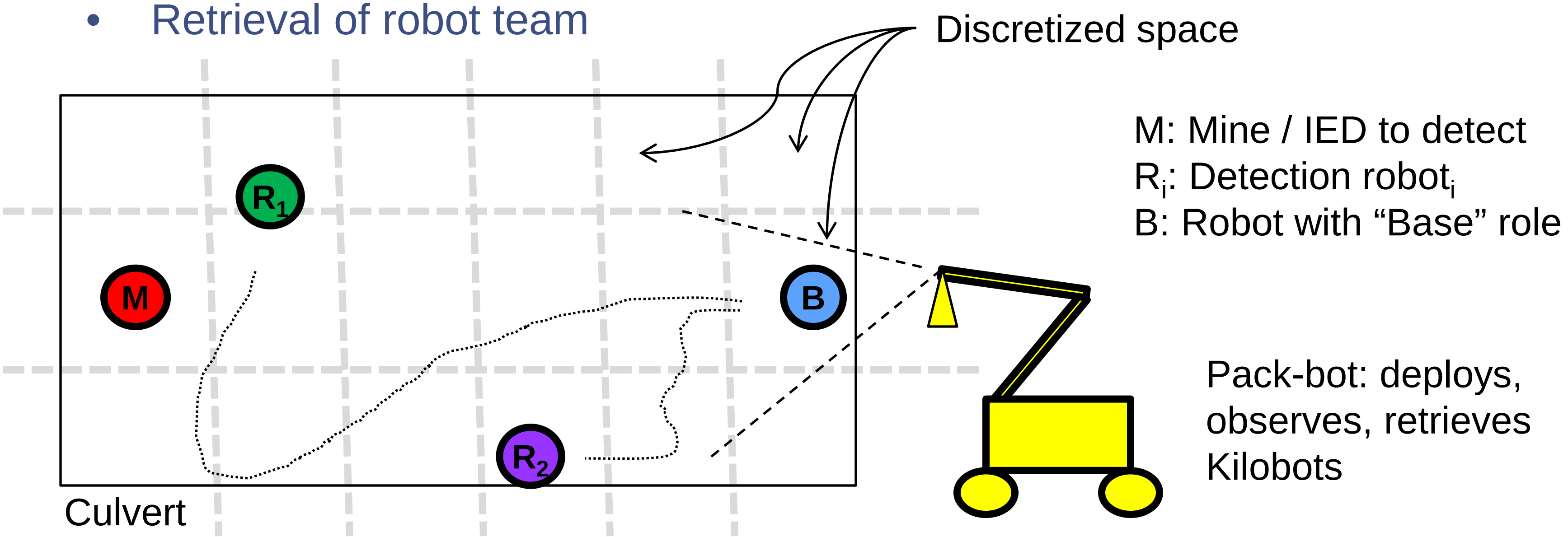
A3: P(detect)=high, P(defuse)=high

Current Validation Experiments

Use individual performance of Kilobots to predict:

- How many robots to deploy for a mission.
- The ordering of robots, in order to maximize:
  - Detection
  - Informing base station
  - Retrieval of robot team





Outline for the Rest of the Presentation

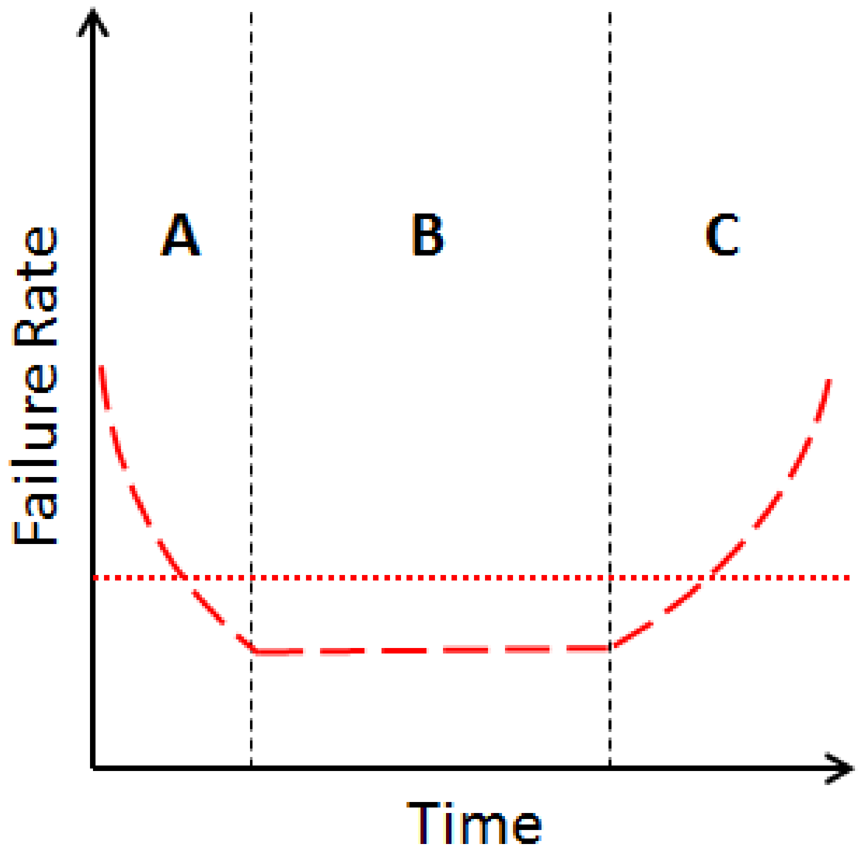
- Robots, Agents, Humans, …
- Other Key Terms & Ways of Thinking
- Technical / Research Approach
- Probabilistic Model Checking
- **Behavioral Reliability Analysis**
- Conclusions and Take-Away Message

Project Team

Developing Coordinated Multi-UGV Reliability Analysis Techniques (MUGVRATS)

- Stephen Blanchette, Jr.
- Kawa Cheung
- John M. Dolan
  - CMU Robotics Institute
  - SAE Reliability Engineering WG Lead
- Joseph A. Giampapa
- David S. Kyle
- John F. Porter, CMU Robotics Institute

Component Reliability Engineering (RE)



- Prior to this project, the only RE model used for evaluating robotic reliability.
- Rather than focus on *failure*, we focus on *performance*.
- How do we derive a characterization for region ***B***?
- Are there analogs for regions ***A*** and ***C***?

Technical Approach

- Understand the atomistic behavioral performance characteristics
  - Biggest challenge: minimizing performance variation
  - Biggest insight: quantifying performance narrows complex root cause search space
- Relate these characteristics to each other
  - Sometimes only a rough characterization is possible
  - Sensor performs best over center of mine, or scanning full area
  - Localization dominates sensing for overall performance
    - Drift is negligible with omni-directional wheels and moderate speed
- Characterize effects of using multiple robots
  - Cyber-physical problem vs. information theoretic one
  - Characterize the roles of robots: complementary, reinforcing, validating
- Predict and validate
  - How to predict derives from above results
  - Validation through experimentation

Understand Behavioral Performance

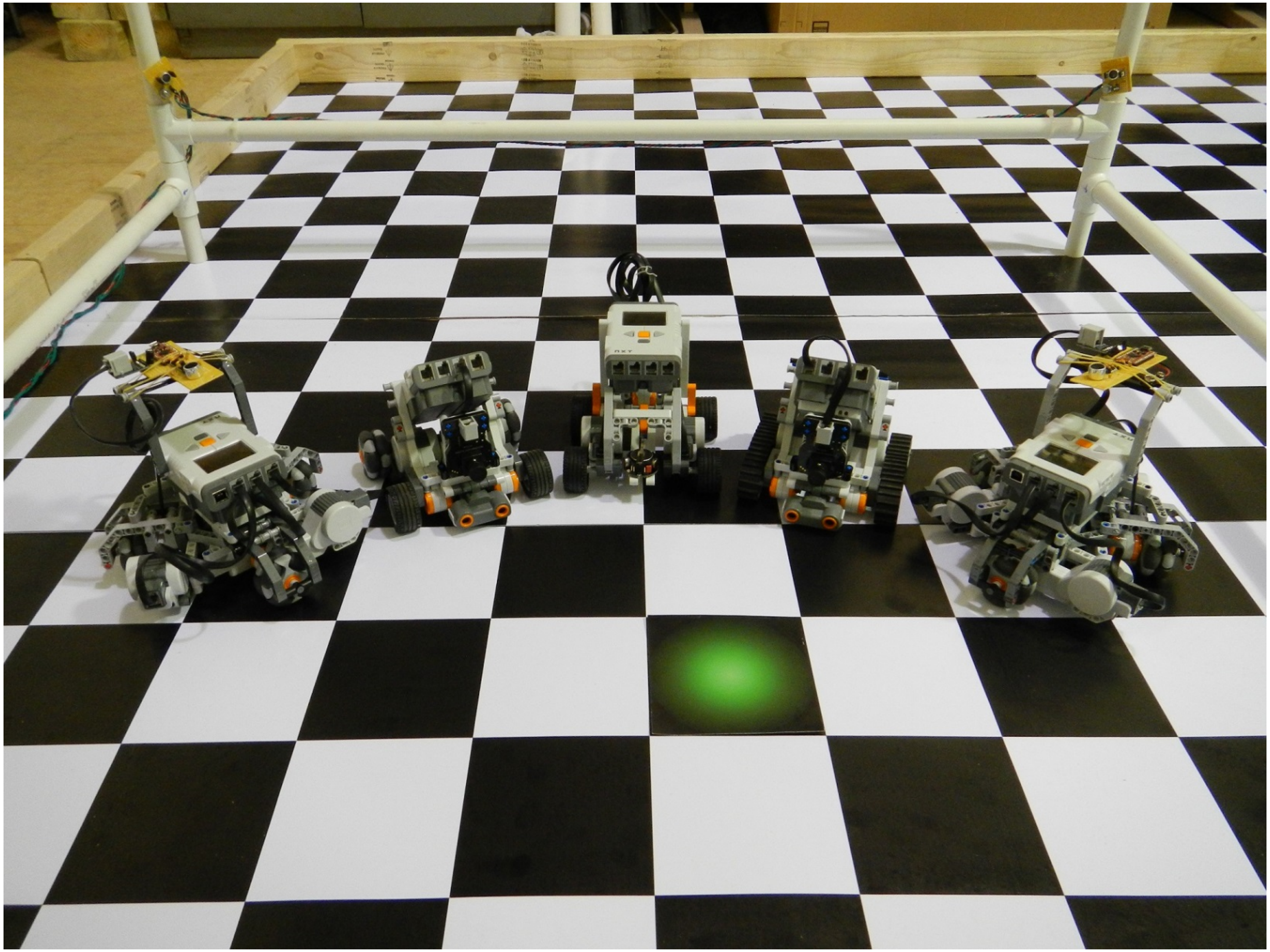
Understand behavioral performance characteristics that are relevant to the mission.

- There are very few: unnecessary features are eliminated from design
- Time to complete mission
- Terrain coverage, depends on localization
  - Is the robot where it thinks that it is?
  - Will there be holes in the scan patterns?
    - This will have a big impact on performance – *PMC insights*
  - Locomotion and motor control have big impact
    - *You can already predict performance from contributions of components*
- Power consumption: negligible for a single experiment run
- Sensor performance
  - Varies per sensor
  - Varies according to “terrain type”
  - Varies according to mine type and depth of mine

Results and Insights Thus Far

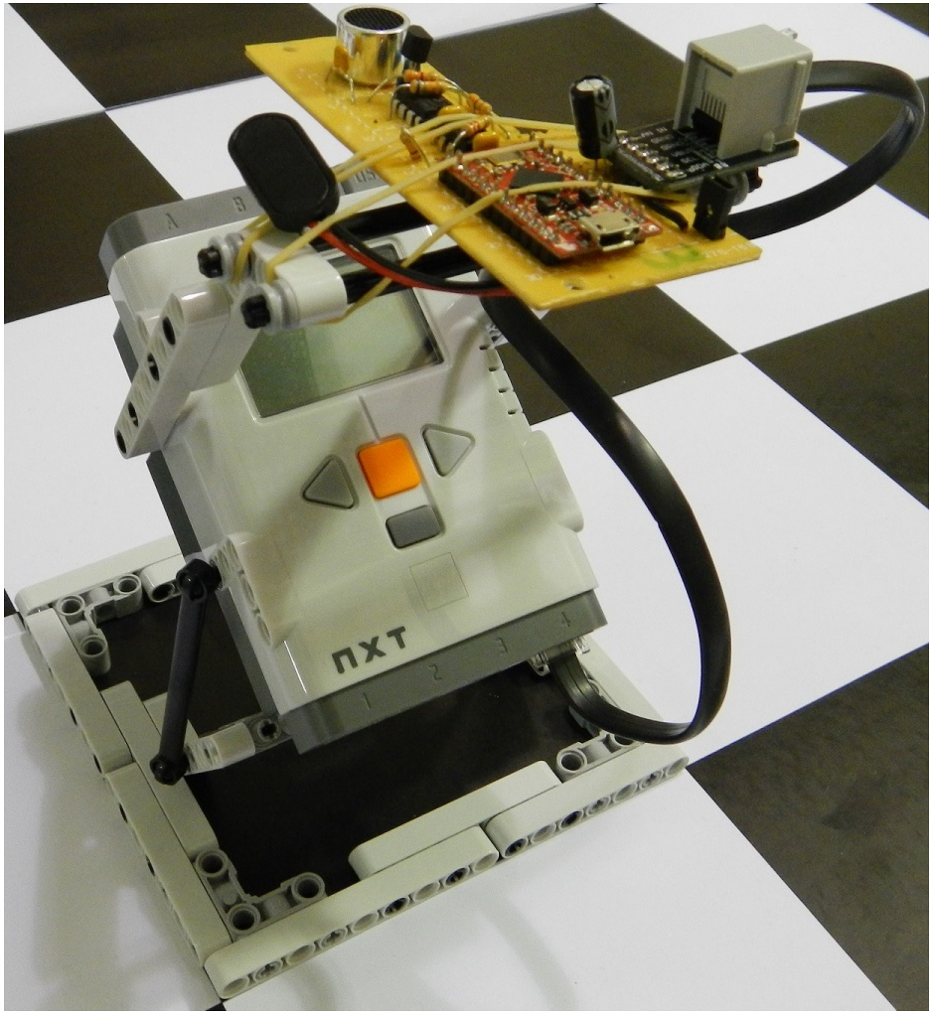
- The technique that best predicts multi-robot coordinated performance:
  - Frequentist likelihood estimates based on features such as:
    - Presence / absence of a mine
    - Individual roles that improve robot localization (e.g. waypoints)
    - Terrain type, because it confuses the sensors
  - This is an information-only model, because robot interactions are only in the information space.
- From Data to Information
  - Analyzing the data allowed us to hypothesize root causes for misbehaviors.
  - Although actual tests to prove root cause are beyond scope of project,
    - The data-driven insights helped us form an accurate “defect model”
    - We could still work with the system while being aware of misbehaviors
      - *This is the reason for assurance in the first place: objective accomplished!*
- Manage Expectations
  - Our performance expectations were revised *once we understood how the metrics behave*.
  - We do not always have the correct intuitions for statistical measures.
  - Robot performance was correct. Our prior expectations for multi-robot performance were not.

Validation Plan: Arena and Robots

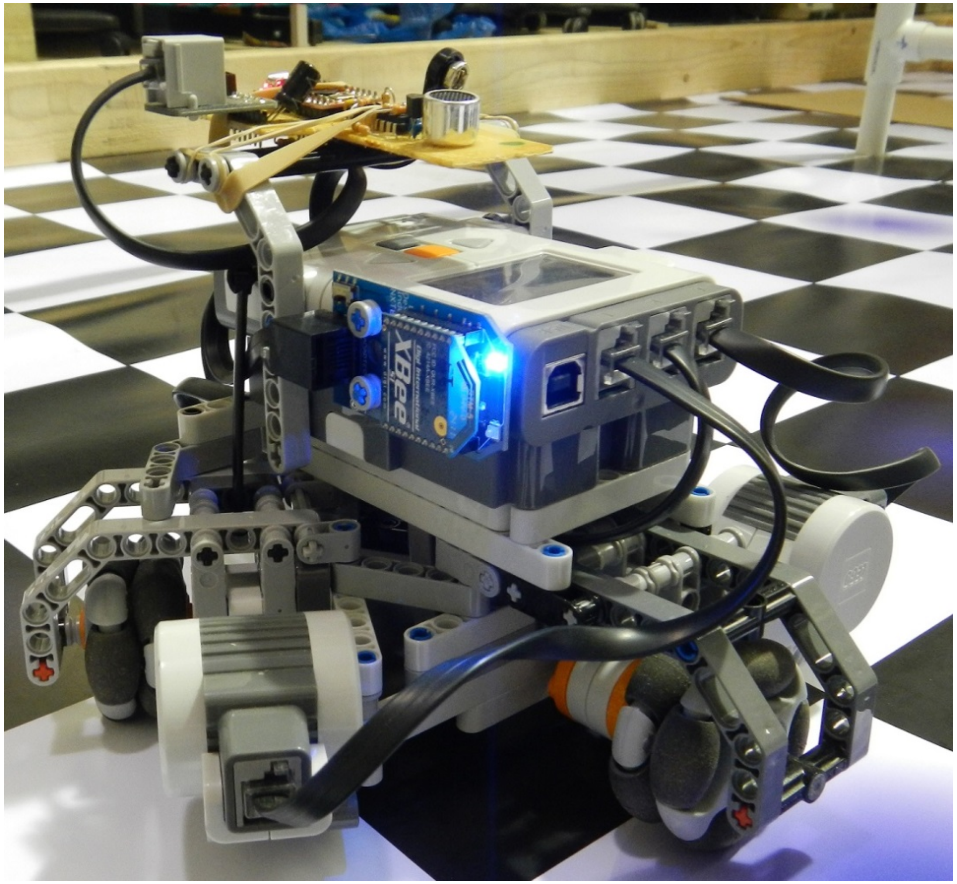


Robot Close-ups

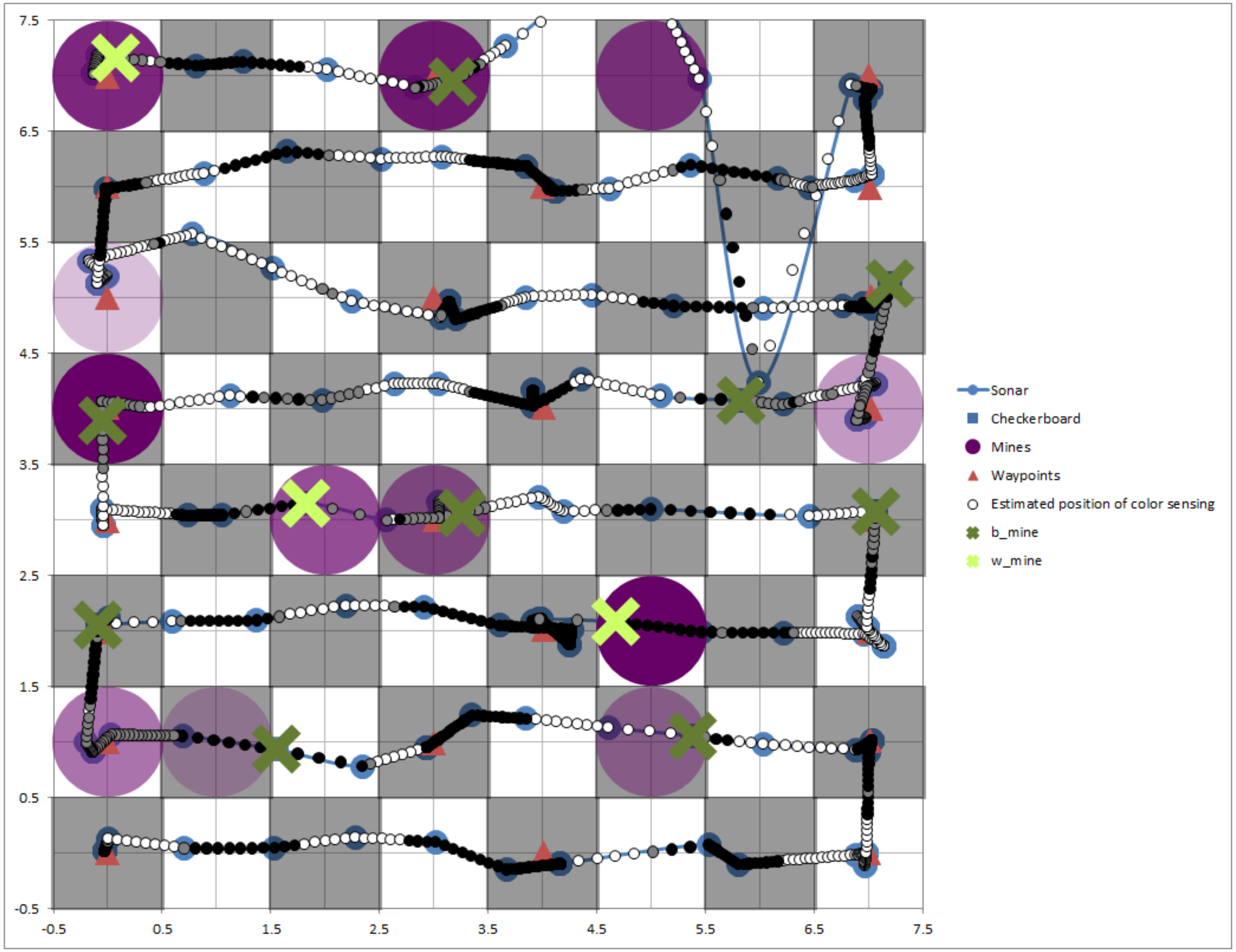
Ultrasound Calibration Rig



Mine Scanning Omni-Bot

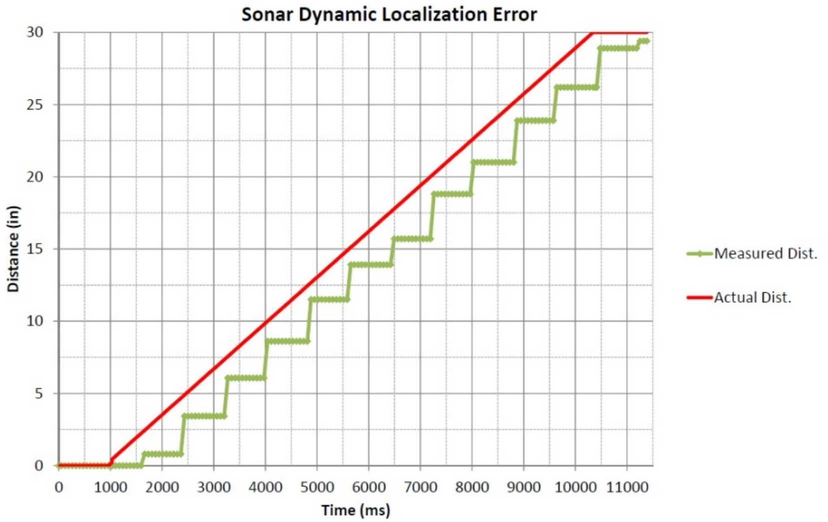


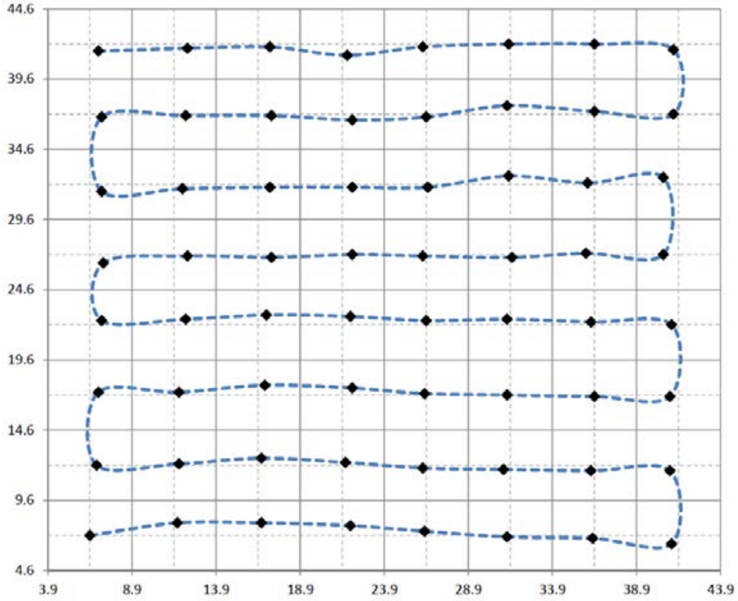
Evaluation of Individual Robot Performance

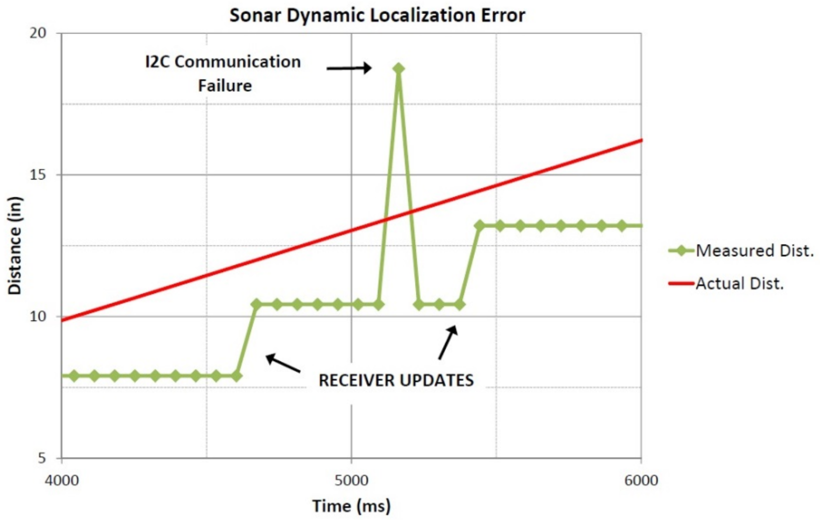


Localization Evaluations

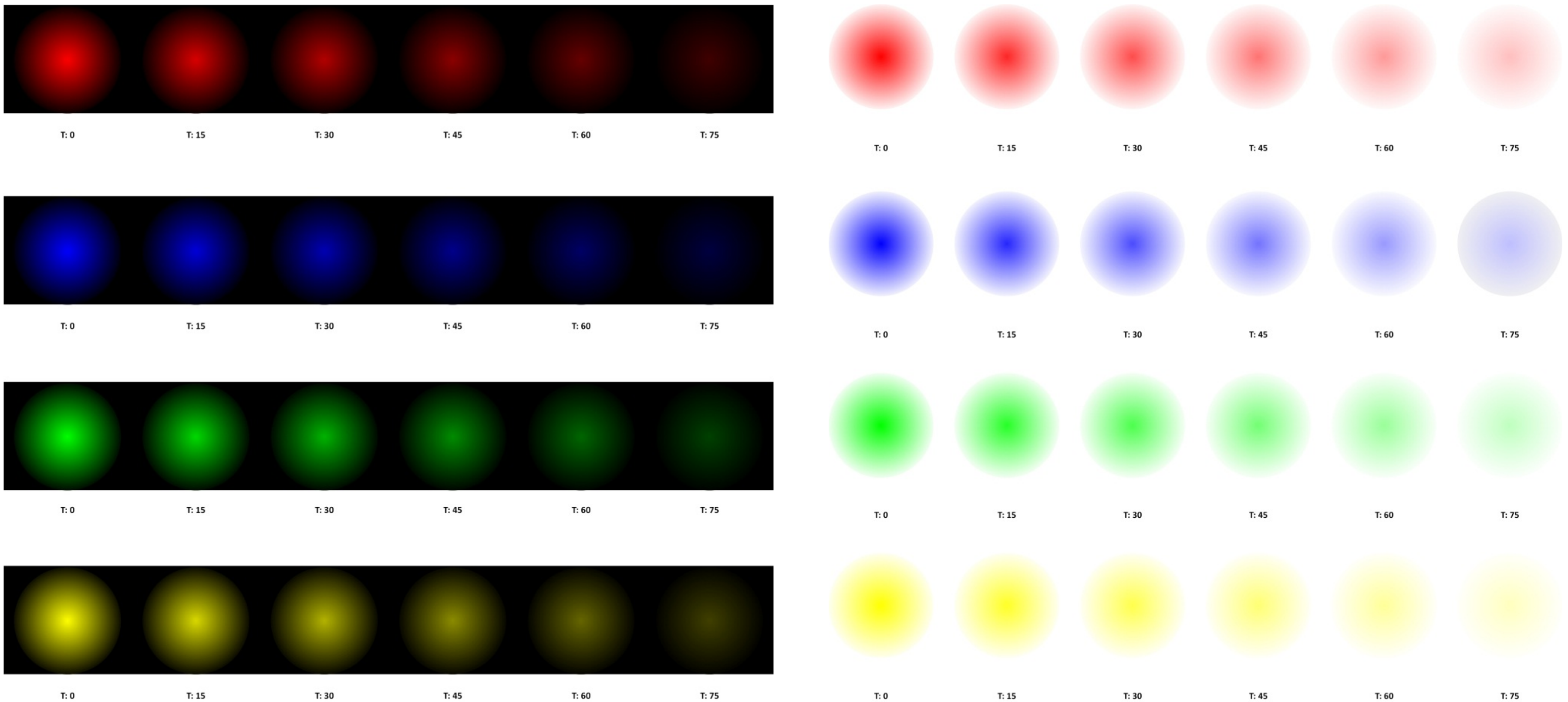
- Bottom-left: static evaluation of ultrasound for each cell
- Top-right: robot localizes self more slowly than actual due to motion
- Bottom-right: sonar localization error due to a communication failure





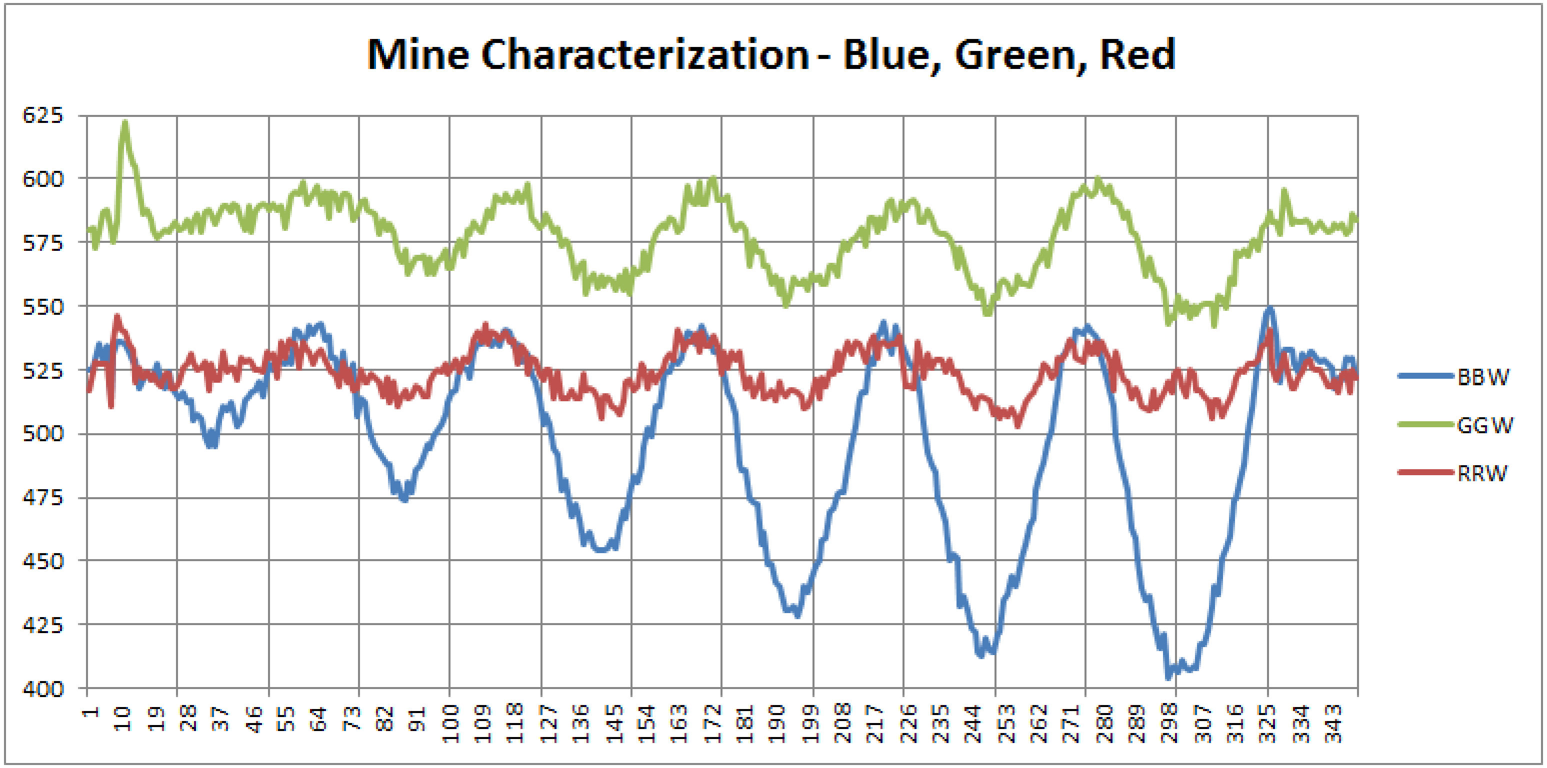


Representation of Mines on Terrain



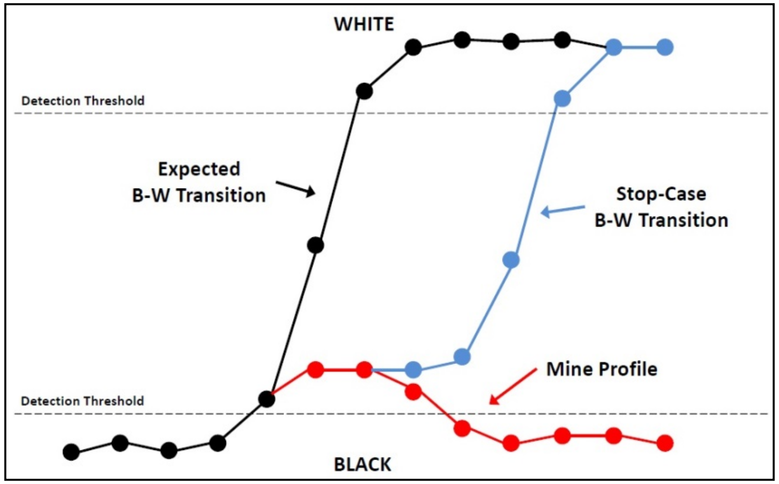
Can you identify what is potentially “wrong” with these test patterns?

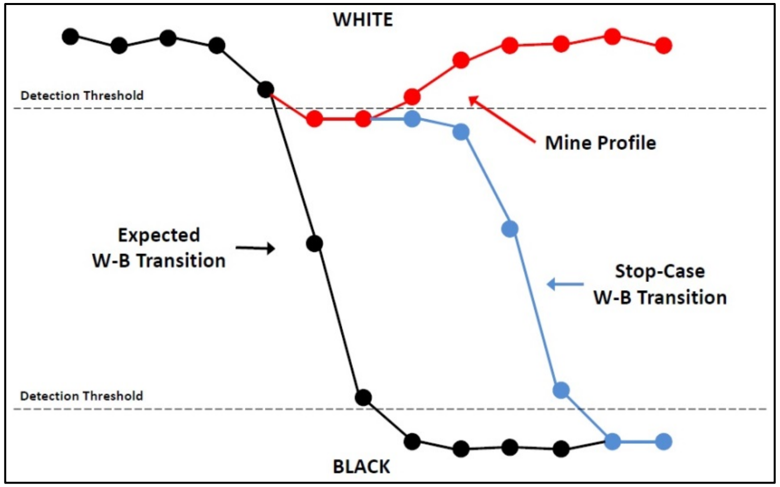
Mean Performance of Robot Sensors

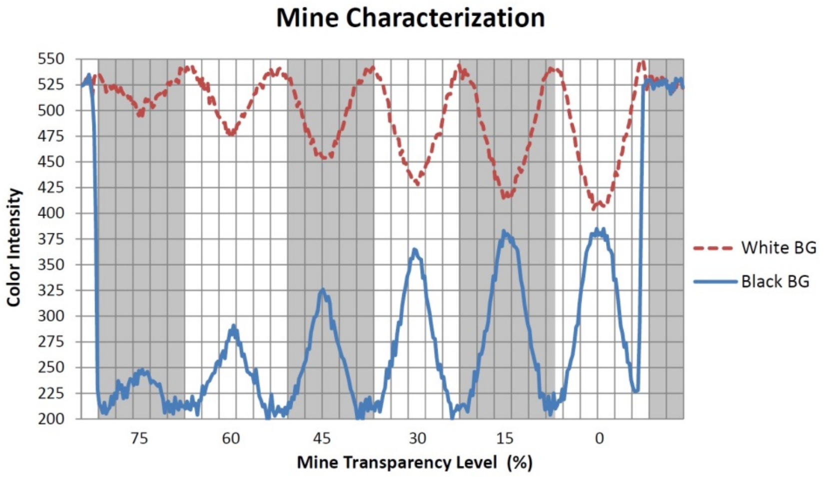


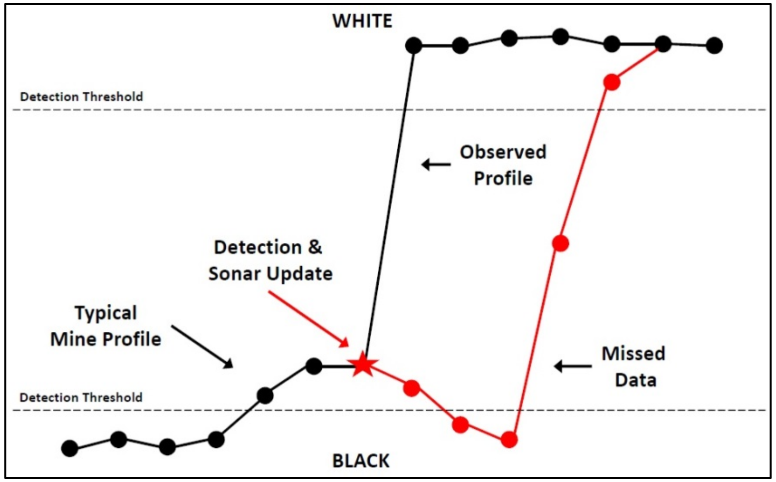
- This characterization was used for sensor selection.
- Given a sensor, you can already predict some aspects of mission performance.

Understanding Root Causes









Additional Information

- S. Chaki, J.M. Dolan, and J.A. Giampapa*, “Toward a Quantitative Method for Assuring Coordinated Autonomy,” in *Autonomous Robots and Multirobot Systems (ARMS 2013)*, at *12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, May 2013. Also published as CMU-RI-TR-13-12.
- S. Chaki and J.A.Giampapa*, “Probabilistic Verification of Coordinated Multi-Robot Missions,” in *First International SPIN Symposium on Model Checking of Software (SPIN13)*, July 2013.
- J.F. Porter, K. Cheung, J.A. Giampapa, and J.M. Dolan, “A Reliability Analysis Technique for Estimating Sequentially Coordinated Multirobot Mission Performance,” in *16th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2013)*, December 2013.

- Talk to me at my information table at this conference.

* Authors ordered alphabetically by surname.

Outline for the Rest of the Presentation

- Robots, Agents, Humans, …
- Other Key Terms & Ways of Thinking
- Technical / Research Approach
- Probabilistic Model Checking
- Behavioral Reliability Analysis
- **Conclusions and Take-Away Message**

Conclusions

- Cost-effective quantifiable assurance techniques for individual and coordinated robots are possible
- Two complementary techniques are being investigated:
  - Probabilistic model checking
  - Reliability analysis
  - Preliminary results are encouraging
- More research is required:
  - To evaluate potential for reuse and shortened assurance processes
  - To account for more coordination phenomena

Take-Away Message

- Competitive advantage for roboticists is grounded in the following:
  - The performance of the autonomous multirobot system
  - Reliable, quantifiable estimates/predictions of system performance
    - Requires assurance cases and arguments
    - Requires assurance arguments to be quantifiable
      - This will force similar quality assurance requirements “upstream” in the supply chain.
- Artifacts that manage expectations for robotic behavioral performance:
  - Assurance cases, logic and evaluation criteria
  - Models to segment performance data by which the evaluations are performed
  - Performance data, itself

Contact Information

Joseph A. Giampapa
Sr. Member of the Technical Staff
Software Solutions Division
Telephone: +1 412-268-6379
Email: garof@sei.cmu.edu

U.S. Mail
Software Engineering Institute
Carnegie Mellon University
4500 Fifth Avenue
Pittsburgh, PA 15213-2612
USA

Web
www.sei.cmu.edu
www.sei.cmu.edu/staff/garof