Trusted Artificial Intelligence Systems Engineering Challenge

WRT1085

Judging Criteria and Prize Update

June 2024

PI: Peter Beling

Deliverables and Judging

- Student teams turn in presentation and white paper at the end of each stage describing:
  - Systems engineering approach,
  - Quantitative results, and
  - Plans for the next stage
- Teams to also submit any software, MBSE models, and other SE artifacts
- Summer submissions are due August 9th
  - Submissions will be shared amongst all teams
- DEVCOM AC and Principal Investigator will serve as judges
  - Optional 20 minute student presentation on afternoon of August 22nd (tentative)
- Three criteria for judging systems solutions to incorporate AI with uncertain performance characteristics:
  - Best Practices (40 percent)
  - Novel Approaches (40 percent)
  - Plans for the next stage of the competition (20 percent)

Awards & Prizes

- $250K in seed funding and prizes – SERC research awards
  - Judges evaluate student deliverables at the end of each stage
  - Contract being modified to provide each team with $10K in seed funding
- Prizes by stage:
  1. Summer 2024: $50K
  2. Fall 2024: $50K
  3. Spring 2025: $70K
- Each team will receive some funding for each stage
- Top teams will receive additional funds
  - Sponsor is currently reviewing proposed prize structure

Confirmed Teams

- Penn State University – Dr. Daryl Farber
- Old Dominion University – Dr. Sachin Shetty
- University of Arizona – Dr. Alejandro Salado
- Purdue University – Dr. Jitesh Panchal
- University of Virginia – Dr. Hunter Moore
- Virginia Tech – Dr. Nathan Lau
- George Washington University – Dr. Zoe Szajnfarber
- Stevens Institute – Dr. Carlo Lipizzi

Back-up Slides

Silverfish Safe Passage: Notional Mission

Safe Passage Mission: Clear mines along defined network

- Mission performance defined as time needed to clear a path from start node to terminus node

Silverfish System:

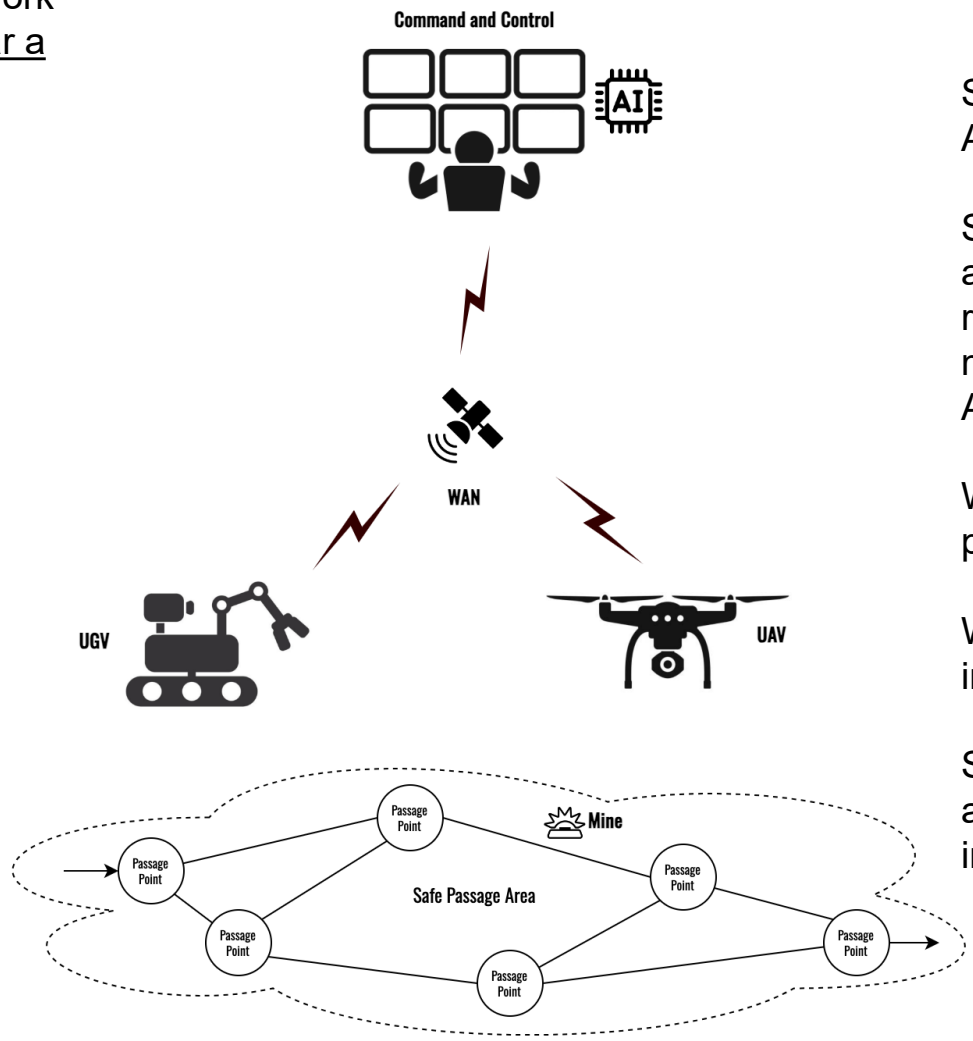
- UGV – mine clearing ground robot scarce resource
- UAV – fast, multi-spectral video collection system

For each link in the network:

- UAV collects video
- AI predicts mine or clear
- AI accuracy is not consistent and can vary with metadata:
  - Ground type
  - Season
  - Topography
  - Lighting/time of day

Command and control center determines UGV path
UGV will positively clear each link it traverses

- 1 hour if mine is encountered
- 20 minutes if mine not encountered



Silverfish System relies on AI to task UGV

Should a human operator assess imagery? Human review of a link takes 30 minutes vs 1 minute for the AI

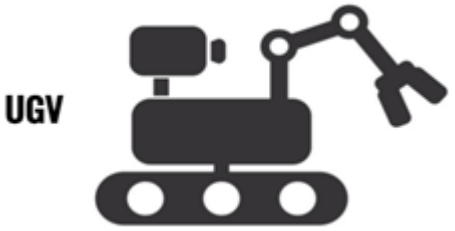
What role should the human play? The AI?

What notions of human trust in AI are important?

Should the system be architected or operated to influence trust?

UGV and UAS Ground Rules and Assumptions

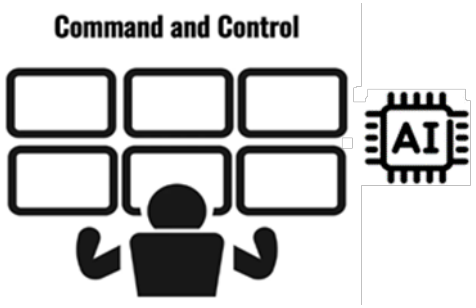
- Unmanned Ground Vehicle (UGV)
  - UGV receives direction from Command and Control on where to go
  - The troops seeking safe passage will always follow the UGV
    - The speed of the UGV limits how fast people can traverse the terrain
  - UGV is 100 percent effective at detecting mines and clearing mines
  - Regardless of the type of terrain, the UGV takes:
    - 20 minutes to traverse the terrain
    - An additional 40 minutes to clear an mines, if a mine is detected
  - UGV has unlimited battery life



- Unmanned Aerial Vehicle (UAV)
  - UAV receives direction from Command and Control on where to go
  - Regardless of type of terrain, UAV takes one minute to capture and transmit imagery to Command Control
  - UAV has unlimited battery life

C2 and WAN Ground Rules and Assumptions

- Command and Control (C2)
  - C2 Operator controls where UAV and UGV go
  - C2 station is equipped with artificial intelligence (AI) that can scan imagery to detect mines
    - Factors impacting how well AI works include terrain, season, time of day, weather, among others
    - One minute needed to scan an image to determine whether a mine is present
    - Unclear how false alarm rate is calibrated regarding false positives and false negatives
  - C2 Operator can have human subject matter expert (SME) scan imagery for mines
    - 30 minutes needed to scan one area to determine whether a mine is present
    - Equal likelihood of false positives and false negatives



- Wide Area Network (WAN)
  - Communications is 100 percent reliable
  - Unlimited bandwidth – no constraints or delays in transmitting imagery or receiving commands

Terrain and Environmental Conditions - Ground Rules & Assumptions

- AI and Human SME ability to detect mines captured during four test events at two locations
  - A variety of terrain conditions across a 10 x 10 grid
    - Rocky, Sandy, Grassy, Wooded, Swampy
  - Different times of day:
    - Location A: 1000 and 2200
    - Location B: 0900 and 2100

Conditions	
currentDateTime	*1000*
currentTemperature	70
currentWindSpeed	5
currentVisibility	0.05
currentPreceiptation	5

AI Performance Table (% Accuracy)

	Row Index									
Column Index	1	2	3	4	5	6	7	8	9	10
1	0.95	0.95	0.95	0.73	0.95	0.95	0.73	0.94	0.95	0.96
2	0.95	0.95	0.95	0.70	0.95	0.73	0.73	0.95	0.96	0.96
3	0.95	0.95	0.94	0.72	0.70	0.72	0.56	0.95	0.96	0.96
4	0.95	0.56	0.95	0.70	0.68	0.56	0.57	0.95	0.95	0.96
5	0.56	0.57	0.68	0.62	0.94	0.57	0.95	0.96	0.94	0.96
6	0.57	0.57	0.68	0.64	0.94	0.57	0.95	0.96	0.95	0.96
7	0.96	0.57	0.70	0.65	0.94	0.56	0.57	0.95	0.95	0.96
8	0.96	0.96	0.68	0.65	0.94	0.56	0.57	0.94	0.96	0.96
9	0.96	0.96	0.69	0.67	0.70	0.56	0.56	0.96	0.96	0.95
10	0.96	0.96	0.72	0.95	0.72	0.56	0.56	0.96	0.96	0.95

Human Performance Table (% Accuracy)

	Row Index									
Column Index	1	2	3	4	5	6	7	8	9	10
1	0.90	0.90	0.90	0.85	0.90	0.90	0.75	0.90	0.90	0.90
2	0.90	0.90	0.90	0.85	0.90	0.75	0.75	0.90	0.90	0.90
3	0.90	0.90	0.90	0.85	0.85	0.75	0.75	0.90	0.90	0.90
4	0.90	0.56	0.90	0.85	0.70	0.75	0.75	0.90	0.90	0.90
5	0.75	0.75	0.75	0.75	0.90	0.75	0.95	0.90	0.90	0.90
6	0.57	0.75	0.75	0.75	0.90	0.75	0.95	0.90	0.90	0.90
7	0.90	0.75	0.85	0.75	0.90	0.75	0.75	0.90	0.90	0.90
8	0.90	0.90	0.75	0.75	0.90	0.75	0.75	0.90	0.90	0.90
9	0.90	0.90	0.75	0.75	0.85	0.75	0.75	0.90	0.90	0.90
10	0.90	0.90	0.85	0.90	0.85	0.75	0.75	0.90	0.90	0.90

Surface Type Table

	Row Index									
Column Index	1	2	3	4	5	6	7	8	9	10
1	Grassy	Grassy	Grassy	Rocky	Sandy	Sandy	Rocky	Sandy	Sandy	Swampy
2	Grassy	Grassy	Grassy	Rocky	Sandy	Rocky	Rocky	Sandy	Swampy	Swampy
3	Grassy	Grassy	Grassy	Rocky	Rocky	Rocky	Wooded	Sandy	Swampy	Swampy
4	Grassy	Wooded	Grassy	Rocky	Rocky	Wooded	Wooded	Grassy	Grassy	Swampy
5	Wooded	Wooded	Rocky	Rocky	Sandy	Wooded	Grassy	Grassy	Grassy	Grassy
6	Wooded	Wooded	Rocky	Rocky	Sandy	Wooded	Grassy	Grassy	Grassy	Grassy
7	Swampy	Wooded	Rocky	Rocky	Sandy	Wooded	Wooded	Grassy	Grassy	Grassy
8	Grassy	Swampy	Rocky	Rocky	Sandy	Wooded	Wooded	Grassy	Grassy	Grassy
9	Swampy	Swampy	Rocky	Rocky	Rocky	Wooded	Wooded	Swampy	Swampy	Grassy
10	Swampy	Swampy	Rocky	Sandy	Rocky	Wooded	Wooded	Swampy	Swampy	Grassy

Stage 1: Summer 2024

- Teams provided with an operational concept and supporting MBSE models for Silverfish Safe Passage
  - https://tsherburne.github.io/silverfish-rsp-report/
- Teams provided with:
  - Performance maps for AI giving per-link accuracy as a function of metadata for several missions
- Develop understanding of how AI performance varies with metadata.
  - For example, teams might characterize envelopes of environmental conditions or other meta data over which AI performance is deemed acceptable.
- Form initial ideas about relationships between model performance and system performance.
- Submissions will consist of a presentation and white paper describing approach and results. These submissions will be made available to all competitors in Stage 2.

Stage 2: Fall 2024

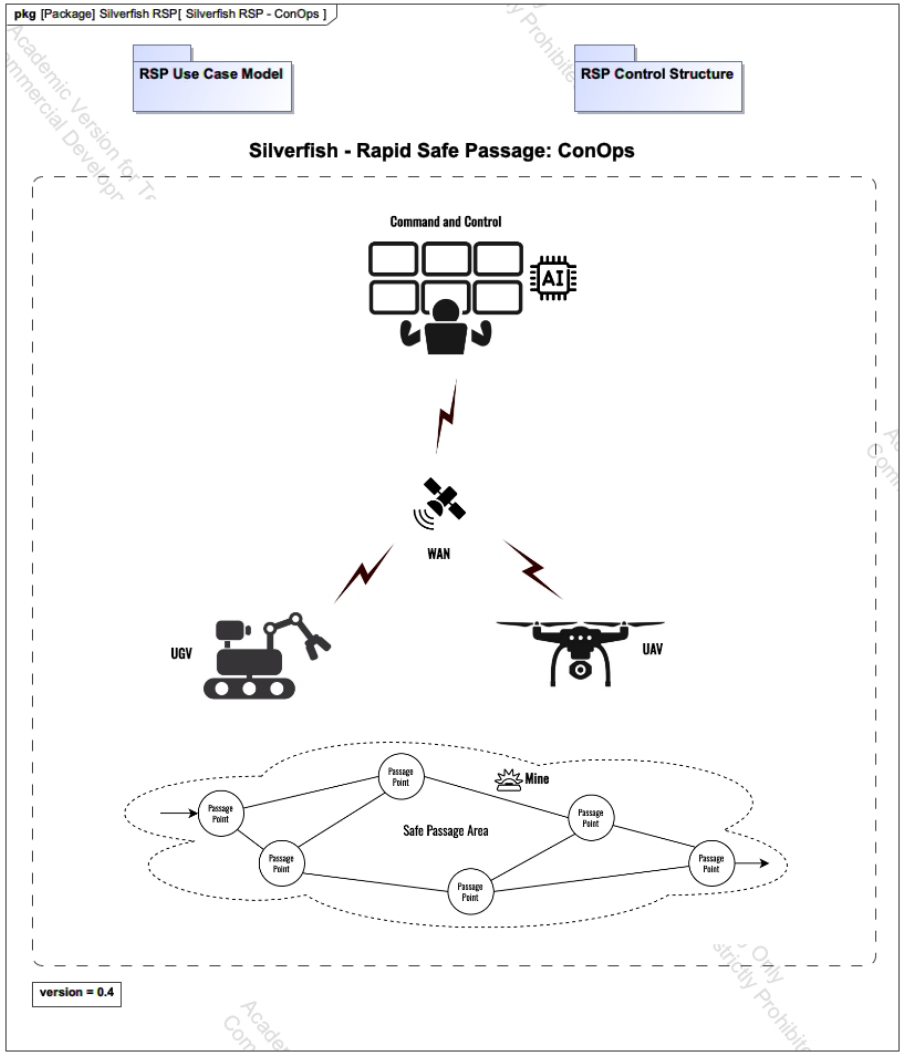
- Stage 2 centers on design of the decision system.
- Designs might consider
  - Architectures for human & machine decision making
  - Methods to increase operator trust
  - Resilience mechanisms based on estimates of model performance
- Teams will be provided operational simulation models for Operation Safe Passage.
- Teams may choose to use AI-based methods for system design, verifications, and validation or to explore modular/open architecture concepts or design for high-level system properties such as safety, security, and trust.
- Submissions will consist of a presentation and white paper describing approach and results. These submissions will be made available to all competitors in Stage 3 for their consideration.
- The top teams may be invited to make conference presentations.

Stage 3: Spring 2025

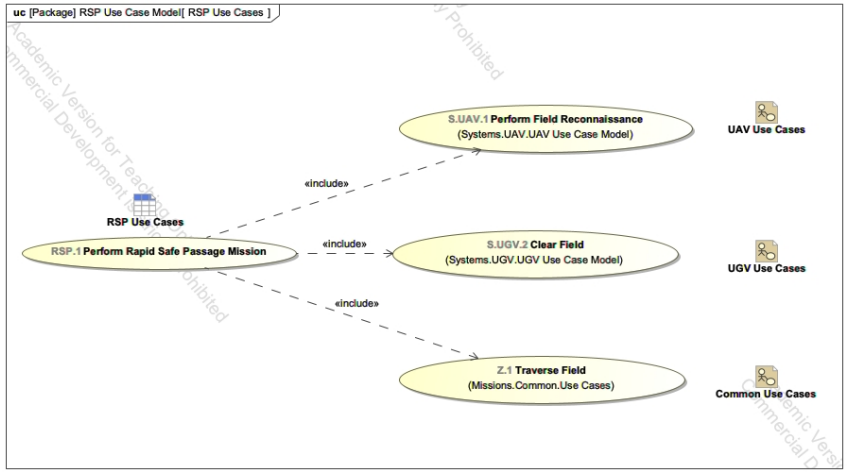
- Stage 3 centers on:
  - Operational simulation of mission scenarios.
  - Extension to lethal scenarios.
- Goals:
  - Exercise system in scenarios not previously seen by the participants.
  - Highlight differences in view
- Scoring will be done on quantitative measures of system performance as well as on the novelty and utility of the SE processes used to generate or validate the design.
- Submissions will consist of an oral presentation, accompanying slides, and final report.
- The top teams may be invited to present at SERC SSRR or the SE4AI/AI4SE workshop.

AI Challenge MBSE Model

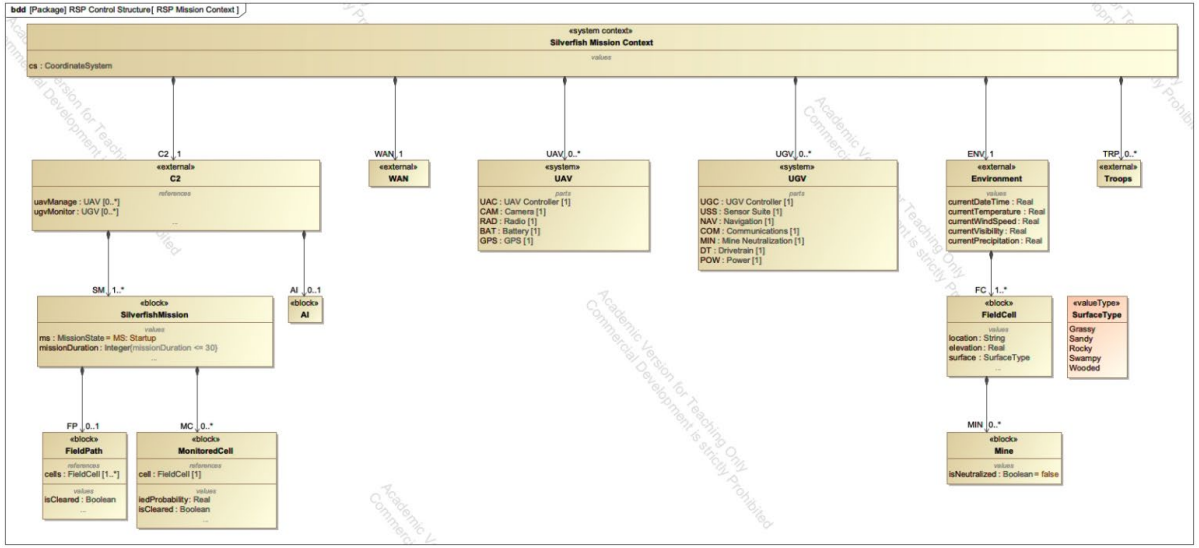
https://github.com/tsherburne/aic



Use Case Model



Mission Context - Block Definition Diagram (BDD)



Thank you

Stay connected with SERC Online:

Email the presenter: Peter Beling

beling@vt.edu

Email the research team: VT National Security Institute

beling@vt.edu