T&E of Autonomous Systems: Challenges and Opportunities of Autonomy Vulnerability T&E
Don Strausberger

Georgia Tech Research Institute
March 2017
Purpose

Drive thought/discussion in the Test Community towards innovative solutions to evaluate vulnerability in the context of autonomous cyber-physical system T&E

Agenda

• Example
• Definitions and Terminology
  • Automation/Autonomy
  • Cyber and Cyber Physical Systems
  • What is autonomy vulnerability?
• CPS Decomposition/Vulnerabilities/Mitigation Solution
• A-CPS Decomposition/Vulnerabilities/Mitigation Solution
• Key Takeaways
• Summary

CPS: Cyber-Physical System
A-CPS: Autonomous Cyber-Physical System
Example - Autonomous-CPS Sea Vessel

Problem – How will autonomy execute the mission, and as necessary, safely navigate COLREGS?

Real-World Scenario

Perceive
Are the other Captains aware of my presence?
What are the vessels current status?

Comprehend
Are they acting in a cooperative, uncooperative, unaware, or adversarial manner?

Project
What is their tactical intent?
What is their strategic intent?
How can I gather more information?
How can I improve the outcome?
What are the options? What is the cost and risk?

Human Captain

Vulnerabilities could exist at all levels of autonomy.

World Model
Is it expected to encounter fishing boats this hour of the day, this time of the year, in this geographic area?
Origins of Autonomous Cyber-Physical Systems

- **Automation** - The system functions with no/little human operator involvement; however, the system performance is **limited to the specific actions** it has been designed to do. Typically these are **well-defined tasks** that have predetermined responses.
  - **Rule-based responses**

- **Autonomy** - The system has a set of **intelligence-based capabilities** that allow it to **respond** to situations that were **not pre-programmed** or anticipated prior to deployment. Autonomous systems have a degree of **self-government** and self-directed behavior.
  - **Decision-based responses**

- **Cyber** - Relating to or involving computers or computer networks...

- **Cyber-Physical Systems**

---

*As defined in the DoD Autonomy Community of Interest (COI) Test and Evaluation Verification and Validation Working Group Technology Investment Strategy 2015-2018

**Cyber-physical systems** (CPS) are smart systems that include engineered interacting networks of **physical and computational components**. These **highly interconnected and integrated** systems provide new functionalities to improve quality of life and enable technological advances in critical areas... *(NIST)*
Different disciplines have developed different terminology in closely related areas.

Terminology does not map one-to-one across disciplines, however they can be associated with the OODA loop to provide a common reference.
A Model of Autonomy

- **Perception**
  - Identification, state, and attributes of relevant objects in a scene

- **Comprehension**
  - Understanding, ordering, and relevance of what was perceived (i.e. “cause and effect”, correlation, etc.)

- **Projection**
  - Assessing possible courses of action and potential outcomes

- **World Model**
  - Past and present beliefs – example: our “world model” of our daily commute allows us to decide if a long line at a traffic light is “normal” or if an event (i.e. accident) has occurred. Our world model grows with experience.

Autonomy vulnerability is the result of as *past and present* external *influences* that *unduly alter* the *current and future* world model, perception, comprehension, and projection abilities of the autonomy, which may result in changes to current and future *decisions*.

---

2 - Authors definition
Vulnerabilities may result from:
1. Altering the information.
2. Altering its temporal, spatial, or geo-temporal attributes.
3. Altering the attributes of relationships of one or more sources of information in the decision-making process.

### Relationships

<table>
<thead>
<tr>
<th>Relationships</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Beliefs and Values</td>
<td>Human-Inspired Beliefs and Values</td>
</tr>
<tr>
<td>2 Unknown entities</td>
<td>Behavior: Cooperative/Hostile,... and Intent: (tactical/strategic)</td>
</tr>
<tr>
<td>3 Peer Relationships</td>
<td>Trusted/untrusted, insightful/detached, ...</td>
</tr>
<tr>
<td>4 External Data Sources</td>
<td>Accurate/inaccurate, trustworthy/ Dated/current Biased/unbiased</td>
</tr>
<tr>
<td>5 Support Team</td>
<td>Uncertain/confident, experienced/inexperienced, engaged/uninvolved</td>
</tr>
</tbody>
</table>
A Cyber Physical System (CPS) Decomposition

People
- Commanders
- Teammates
- Operators

Data

Cyber-Physical System (CPS)
- Automation
- Cyber
- Autonomy
- Cyber-Physical Systems

World

Sensors

Automation/Complex Cyber Subsystems
Vulnerability Mitigation in a CPS

Cyber-Physical System (CPS)

- Sensors

- Automation/Complex Cyber Subsystems

- Education, Training, and Experiences Specialization
- Systems Vulnerability Specialization
- Communications Vulnerability Specialization
- Cyber Vulnerability Specialization
- Sensor Vulnerability Specialization

People
- Commanders
- Teammates
- Operators

Data
### Addressing CPS Vulnerabilities

**Solved today with two complementary paths, each rooted in different foundations.**

<table>
<thead>
<tr>
<th>Problem</th>
<th>Scenario with Technical Solution</th>
<th>Mission (“Commanders Intent”) with Warfighter Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope</td>
<td>Domain/task specific sensing, perception, and decision making</td>
<td>Warfighter-mission perception, comprehension, projection, and decision making</td>
</tr>
<tr>
<td>Solution</td>
<td>Solved with evolving T&amp;E Capabilities, Design of Experiments to meet complex SoS demands.</td>
<td>Solved with established warfighter training, and evolving T&amp;E-enabled infrastructure (JMETC, etc.), and operationally relevant test and training environments.</td>
</tr>
<tr>
<td>Foundation</td>
<td>“Scientific Method” Laws of physics, Statistics, ...</td>
<td>US Warfighter Values and Beliefs, Attributes of information: Trust &amp; Confidence; Attributes of others (humans): Intent, Trust, and Confidence; Expectations of self/of others</td>
</tr>
</tbody>
</table>
Vulnerability Mitigation in an A-CPS

People
- Commanders
- Teammates
- Operators

Data

Autonomous Cyber-Physical System (A-CPS)
- Perceive
- Comprehend
- Project
- Decide

Sensors

Automation/Complex Cyber Subsystems

World

Cyber-Physical Platform (CPS)

World Model

"???"

- Systems Vulnerability Specialization
- Communications Vulnerability Specialization
- Cyber Vulnerability Specialization
- Sensor Vulnerability Specialization
While the problem and scope do not change, a *new foundation* must be established and an appropriate *solution* must be formed.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Scenario with Technical Solution</th>
<th>Mission (&quot;Commanders Intent&quot;) with Warfighter Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope</td>
<td>Domain/task specific sensing, perception, and decision making</td>
<td>Warfighter-mission perception, comprehension, projection, and decision making</td>
</tr>
<tr>
<td>Solution</td>
<td>Solved with evolving T&amp;E Capabilities, Design of Experiments to meet complex SoS demands.</td>
<td>Do established warfighting training solutions apply? i.e. do we seek to train our autonomy and establish standardized evaluations for it to “pass” and become “qualified”? Do technical solutions “port” over? i.e. Should/can we define “autonomy attack surfaces”?</td>
</tr>
<tr>
<td>Foundation</td>
<td>Foundation: “Scientific Method” Laws of physics, Statistics, ...</td>
<td>Does the autonomy have Values and Beliefs? What attributes does the autonomy give to unknown entities, to peers, to data sources, to support team? How are these attributes formed?</td>
</tr>
</tbody>
</table>
Key Takeaways

- The human operator/controller decision-making in today’s Cyber Physical System exists because of his/her foundation of:
  - Human-inspired US warfighter Values and Beliefs
  - Human-inspired abilities to attribute trust and confidence in information
  - Human-inspired abilities to attribute intent, trust and confidence in others
  - Human-inspired abilities to establish expectations of self and others

- Today’s vulnerabilities of the decision making process are assessed and mitigated based upon this human-inspired foundation

- This foundation naturally enables today’s decomposition into complimentary “technical” solutions and “training” solutions
Key Takeaways

- The autonomy decision-making in tomorrow’s Autonomous Cyber Physical System has a very different foundation:
  - *Machine-based* Warfighter Values and Beliefs
  - *Machine-based* abilities to attribute trust and confidence in information
  - *Machine-based* abilities to attribute intent, trust and confidence in others
  - *Machine-based* abilities to establish expectations of self and others

- Tomorrow’s vulnerabilities of the decision making process will need to be assessed and mitigated based upon these *machine-based abilities*

- This (significantly different) foundation limits today’s “training-based” solutions from solving tomorrow’s autonomy decision-making vulnerability
Assessing vulnerability of the autonomy in an Autonomous-Cyber Physical System is a significant challenge rooted in foundational changes in transferring decision making from a human to a machine.

Solutions may come from a combination of the following:

- Carrying forward and applying established (communication, cyber, sensor, etc.) vulnerability domain T&E toolsets and methodologies to the complex Cyber-Physical System/SoS – “leverage”

- Expanding established vulnerability domain T&E methodologies such as attack surface identification, attack vector generation, etc., to evaluate autonomous decision making - “build”

- Adaptations to today’s training-centric solution “change”
Questions/Comments?

- Contact Information
  - donald.strausberger@gtri.gatech.edu