THE ROAD TOWARDS AUTONOMOUS CYBERSECURITY: REMEDIES FOR SIMULATION ENVIRONMENTS

MARTIN DRAŠAR, ÁDÁM RUMAN, PAVEL ČELEDÁ, SHANCHIEH JAY YANG
DRASAR@ICS.MUNI.CZ
SECAI 2023
HOW TO GET TO AUTONOMOUS CYBERSECURITY?

• Despite all the promises of AI, we are not getting anywhere with autonomy
• Numerous reasons:
  • Domain complexity
  • Insufficient training datasets
  • Insufficient tooling
• This presentation addresses those reasons through the prism of training environments
STATE OF THE ART

• Training environments are under-researched and under-developed
• Generic solutions cannot be used, they do not capture the complexity
• Cybersecurity solutions are either too abstract, or too specific
• Narrow scope of tooling

• No real push for creating deployable solutions
SIMULATION ENVIRONMENTS

- Often build as means to an end
- No solid theoretical foundation
- In effect, different environments are incomparable
  - At least until today…
ASSESSMENT FRAMEWORK FOR AUTONOMOUS CYBER AGENT SIMULATION

- Simulations cyber terrain ontology
- Actor evaluation framework
- Comprehensiveness and concreteness measurement
SIMULATIONS CYBER TERRAIN ONTOLOGY

- **Topology Plane**: Physical topology of the infrastructure
- **Logic Plane**: Functionality of simulation
- **Meta-communication Plane**: Inter-plane signaling
- **Supervision/Observation Plane**: Observation spaces and timing
- **Actor Plane**: State-changing entities
ACTOR EVALUATION FRAMEWORK

- **Based on COI framework**
  - **Intent**: existential goal of actor
  - **Opportunities**: domain of events that can be invoked by actors
  - **Capabilities**: predicates limiting actor’s opportunities
  - **Preferences**: prioritization based on secondary intents
  - **Sophistication**: cost and risk assignment of taking specific opportunities

- **Four generic actor types**: Adversaries, Defenders, Benign Participants, Fates
COMPREHENSIVENESS AND CONCRETENESS MEASUREMENT

- **Based on MITRE’s metrics**

- **Perspectives**: Attack vectors, attack actions, adversary characteristics, defender actions, technical architecture, technical vulnerabilities

- **Concreteness**: Abstract, notional, representative, fully realized
- **Comprehensiveness**: Fragmentary, partially specified, fully specified
ASSESSMENT OF DEPLOYABILITY

• We assert that to create deployable solutions, the training environment has to:
  • Approach minimal abstraction
  • Provide actionable descriptions of the terrain, users, vulnerabilities, etc.
  • Be dynamic and able to evolve
  • Be concrete and comprehensive

• To this end we analyzed the following environments:
  • Yawning Titan, CyberBattleSim, CybORG, CYST, and NaSimEmu (not in the paper)
COMPARISON OF CYBORG AND CYST

• According to assessment framework the two most sophisticated tools
• We made a qualitative evaluation based on 2nd CAGE challenge and similar custom scenario for CYST
• We evaluated strong points from the point of view of a developer of autonomous cybersecurity system.
CAGE challenge

CYST scenario
• **Both environments gravitate to similar goals and use similar approaches**

• **CybORG is more readily usable at the expense of advance features**

• **CybORG dropped emulation support, so it is unusable in the future**

<table>
<thead>
<tr>
<th>CYST</th>
<th>CybORG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infrastructure &amp; Logic</strong></td>
<td><strong>Service and OS knowledge base.</strong></td>
</tr>
<tr>
<td>Network traffic shaping.</td>
<td>Modeling OS.</td>
</tr>
<tr>
<td>Modeling the traffic.</td>
<td>Support for complex authentication and Host level information down to PID and authorization.</td>
</tr>
<tr>
<td>Supervision, Actors &amp; Agents</td>
<td>Modeling OS.</td>
</tr>
<tr>
<td>Non-singular action handling.</td>
<td>Transaction support for faster training.</td>
</tr>
<tr>
<td>Agent-agent interaction in addition to agent-environment.</td>
<td>Ready wrappers and interfaces for OpenAI.</td>
</tr>
<tr>
<td><strong>External &amp; Miscellaneous</strong></td>
<td></td>
</tr>
<tr>
<td>Strong focus on deployability.</td>
<td>Maximizing extensibility, stand-alone packages, usable as a library, and plugin support.</td>
</tr>
<tr>
<td>Integration with outside running services.</td>
<td>Human-machine interface.</td>
</tr>
</tbody>
</table>
• Multi-agent discrete-event simulation framework tailored for cybersecurity
• Highly extensible and flexible (action spaces, observation spaces, …)
• Supports transformation of simulation artifacts into flows, packet traces, etc.
• Enables integration of simulation and emulation (IDS in the loop)

• HTTP://MUNI.CZ/GO/CYST
• HTTPS://GITLAB.ICS.MUNI.CZ/CRYTON/BEAST-DEMO
AI-DOJO

- Research project to create a platform for development of autonomous cybersecurity systems
- Integration of simulation and emulation
- Library of agents with different behavior (attacker, defenders, users)
- Automated generation of realistic cybersecurity scenarios to support learning

https://muni.cz/go/ai-dojo
AICA-IWG

- **Follow-up to NATO IST-152 tasked with specification of Reference Architecture for Autonomous Cyberdefense Systems**
- **Working group focused on furthering the development of Autonomous Cybersecurity Systems**
- **Academia, Industry, Defense**

- [HTTPS://WWW.AICA-IWG.ORG/](HTTPS://WWW.AICA-IWG.ORG/)
ADDENDUM: NASIMEMU ASSESSMENT
Abstraction Level: High

Topology:
- Dynamic Changes: Allowed
- Representation: Custom Data Structure

Logic:
- Network:
  - Rule Direction: Bidirectional
  - Rule Granularity: Per protocol, for subnets
  - Additional Capabilities: None
- Hosts:
  - OS: Available, high-level tags
  - Software: Process
  - Software Properties: Versions using tags
- Users:
  - Account Granularity: Not supported
  - Credentials: Not supported
  - Authorizations: Only the level of control over a host
  - Remote Access Control: Not supported
  - Local Access Control: User privileges
- Weaknesses:
  - Reliability: High
  - Representation: Exploitable vulnerabilities
  - Applicability Guard: Service name
  - Additional Action Attributes: Cost, probability of success

Meta Communication:
- Event Invocation: Supervision Intervention
- Event Propagation: Supervision Intervention

Supervision/Observation:
- Observation Space: Provided
- Timing: Sequential
- Reward Computation: Provided
- Multi-Agent Support: Unknown

Adversarial COPS:
- Intent: Finding a pre-defined loot
- Opportunities: 8 Actions (Exploit, Privilege escalation, ServiceScan, ProcessScan, Terminal Action)
- Capabilities: Accessible hosts (via control level), vulnerabilities
- Preference: Customizable via reward computation
- Sophistication: Agent dependent

Defender COPS:
- Defenders not available

Concrete:
- Adversary Characteristics: Abstract
- Attack Vectors: Notional
- Attack Actions: Abstract
- Defender Actions: Unavailable
- Technical Architecture: Notional
- Technical Vulnerabilities: Representative

Comprehensiveness:
- Adversary Characteristics: Fragmentary
- Attack Vectors: Partially specified
- Attack Actions: Partially specified
- Defender Actions: Unavailable
- Technical Architecture: Fragmentary
- Technical Vulnerabilities: Partially specified