STPA APPLIED TO A NEURAL NETWORK-CONTROLLED AIRCRAFT

RYAN BOWERS – 40TH FLIGHT TEST SQUADRON
DR. JOHN THOMAS – MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Overview

- 40th FLTS: flight test for AI-enabled autonomous aircraft
- July 2023: First flight test of a group-5 UAV flown by machine learning agents
- Agents trained using deep reinforcement learning
- Applied STPA before flight test
Deep Reinforcement Learning Agents

- Computes actions based on environment observations

**Environment state**
- UAV state
- Safety limits
- External entities

**Flight controls**
- $n_z$, roll rate, throttle
System Considerations for Safety

- ML agents can be difficult to explain

- Agents trained in simulation, then transitioned to real life

- UAV and agents developed under completely separate programs before integrating
Three-Pronged Flight Test Safety Approach

(1) UAV Mechanisms
**Envelope trips:** Disables agent if speed/altitude limits exceeded

**Command Limiters:** Agent control inputs are clipped to stay within min and max bounds.

(2) Autonomy Mechanisms
**Simulation Training:** Agents were trained to stay within limits.

**Redundant envelope trips:** Agent disables itself if limits exceeded.

(3) Test Procedures
**Manual Disable:** Remote pilot can disable agent at anytime.

**Abort Limits:** Manually disable if any limits exceeded.

**Briefing Items:** Team briefed on possible unsafe agent behavior.
STPA: System Theoretic Process Analysis

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

What can go wrong in flight test?
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

Remote Operator (RO)

Autonomous Controller (AC)

Physical Aircraft

Control Actions

Feedback

Decision Making

Process Model

Not provided causes hazard
Providing causes hazard
Too early, too late, out of order
Stopped too soon, applied too long

UCA1  UCA2  UCA3  UCA4

Waveoff (abort landing) when ...

STPA: Identify Unsafe Control Actions (UCAs)
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

Remote Operator (RO)
- Decision Making
- Process Model

Autonomous Controller (AC)
- Decision Making
- Process Model

Physical Aircraft

Control Actions

Feedback

Env.

STPA: Identify Unsafe Control Actions (UCAs)

<table>
<thead>
<tr>
<th>UCA1</th>
<th>UCA2</th>
<th>UCA3</th>
<th>UCA4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not provided causes hazard</td>
<td>Providing causes hazard</td>
<td>Too early, too late, out of order</td>
<td>Stopped too soon, applied too long</td>
</tr>
</tbody>
</table>

UCA2: AC provides Waveoff when fuel too low
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

UCA2: AC provides Waveoff when fuel too low

Why would this happen?
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

Remote Operator (RO)

Autonomous Controller (AC)

Decision Making

Process Model

Control Actions

Feedback

Physical Aircraft

Env.

PM: AC believes crosswind too high to land

DM: If crosswind too high, immediately Waveoff

UCA2: AC provides Waveoff when fuel too low

E: Sideslip at altitude is high. (no crosswind info)
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

DM: If **crosswind** too high, immediately Waveoff.

UCA2: AC provides **Waveoff** when fuel too low.

PM: AC believes **crosswind** too high to land.

E: **Sideslip** at altitude is high. (no crosswind info)
STPA applied to humans (FTEs, Pilots, etc.)
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

Remote Operator (RO)
- Decision Making
- Process Model

Autonomous Controller (AC)
- Decision Making
- Process Model

Control Actions

Feedback

STPA: Identify Unsafe Control Actions (UCAs)

Force Landing when ...

AC provides Waveoff when fuel too low

Providing causes hazard
Not provided causes hazard
Too early, too late, out of order
Stopped too soon, applied too long

Env.
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

UCA: RO does not provide Force Landing Cmd when fuel is too low

Remote Operator (RO)
- Decision Making
- Process Model

Control Actions

Feedback

Autonomous Controller (AC)
- Decision Making
- Process Model

Control Actions

Feedback

Physical Aircraft

Receives report that crosswind is ok on ground

PM: RO believes aircraft will land (fuel low, crosswind ok, etc.)

FB: Fuel is Low

FB: No crosswind (not included in AC feedback to RO)

AC provides Waveoff when fuel too low

Env.

(John Thomas, 2021)
STPA Step 3: Identify UCAs

UCA: RO does not provide Force Landing Cmd when fuel is too low

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

Remote Operator (RO)

- Decision Making
- Process Model

Autonomous Controller (AC)

- Decision Making
- Process Model

Physical Aircraft

Control Actions

Feedback

AC provides Waveoff when fuel too low

Receives report that crosswind is ok on ground

PM: RO believes aircraft will land (fuel low, crosswind ok, etc.)

FB: Fuel is Low

FB: No crosswind (not included in AC feedback to RO)
Results

- Gaps in procedure

- Re-request landing (will abort again)
- Delayed override cmd while fuel low

- SW algorithm to prematurely abort before lower alt. meas.
- Crosswind vs. low fuel priority: bad permanence assumption

- No warning before landing abort
- Misleading landing confirmation (did not consider crosswind)
- No crosswind measurement (considered redundant)

- Missing ground crosswind feedback
- SW assumption that sideslip at X ft = crosswind at landing site

Remote Operator (RO)

- Decision Making
- Process Model

Autonomous Controller (AC)

- Decision Making
- Process Model

Physical Aircraft

Env.

• Inconsistency: Operator MM doesn’t match Software PM / CA
STPA Application

Session 1: 4-day training + application (UAV only)
Session 2: 5-day training + application (UAV + AI)
Session 3: 2-day application (UAV + more AI detail)

Scope of analysis:
- Focus on flight test ops rather than internal system design
- Black-box AI – could do anything at any time
- Can the Autonomy Safety Sandbox handle all situations?
System Control Structure

- **Autonomy status**
- **UAV status**
- **Autonomy commands**
- **Autonomy in control Y/N**

**Remote Pilot**
- Voice comm: Weather, nearby aircraft

**Chase Pilot**
- Voice comm: Weather, nearby aircraft

**Test Conductor**
- Start next test point
- Terminate/abort
- Initiate recovery
- Deconfliction (weather, aircraft)

**Autonomy Operator**
- Load autonomy task
- Disable autonomy

**Agent**
- Roll rate command
- Nz command
- Throttle command
- Request control
- Self-Disable

- Consent for control
- UAV state and limit data

**UAV Safety Mechanisms**
- Airspeed
- Altitude
- Nz
- Attitude
- Limit autonomy command magnitude

**Auto-Disable**
- Disable autonomy

**Command Limiters**
- Achieve zero roll
- Achieve zero pitch
- Stabilize at current speed
- Give UAV control to remote pilot

**Transition Maneuver Automation**
- Airspeed
- Altitude
- Nz
- Current roll, pitch

**Host UAV Flight Computer**
- Control surface deflections
- Control surfaces – Roll, Pitch, Throttle

**Legend**
- Control action
- Feedback
- Human
- System

** Humans in directive role**

**Human Operators**
- Enable autonomy mode
- Disable autonomy mode
- Fly UAV
- Initiate recovery

**Major Software Systems**

**Coordination through voice comm**
Finding 1: Limitations of Command Limiters

- Command limiters not complex enough to prevent some unsafe/inefficient commands
- No prevention of unsafe input **combinations**
- No awareness of time history – divergent **oscillatory** control inputs possible
- Recommendation: implement mechanism to prevent unsafe **maneuvers**
Finding 2: Inflexible UAV Auto-Disable Mechanism

- Auto-Disable altitude/airspeed bounds could not be easily modified
- Could not test Auto-Disable mechanism without assaulting the real limits
- Recommendations:
  - Make limit enforcement mechanisms flexible
  - Early tester involvement in system design
Finding 3: Incomplete Feedback from Autonomy to Remote Pilot

- Remote pilot had no direct indication of agent’s status or actions.

- "Autonomy mode" did not always mean the agent was in control.

- Recommendation: Provide unambiguous indication of agent status to the remote pilot.

Diagram:

- Remote Pilot
  - Fly UAV
  - UAV state

- Agent
  - Autonomy mode status (on/off)
  - Roll rate command
  - Nz command
  - Throttle command
  - Request control
  - Consent for control
  - UAV state and limit data

- Host UAV Flight Computer

Does not explicitly mean that Autonomy is active
Conclusions – Autonomy Safety Sandbox

- Three-pronged safety framework was effective but imperfect

- UAV safety mechanisms would not prevent all likely concerns

- Can mitigate those concerns by adding/modifying test procedures, but that tends to be heavy handed

- Some issues required band-aids because system design was fixed – recommend STPA during design
Conclusions – Use of STPA

- STPA was effective in identifying new test hazards and gaps
- Does not need to be the only method – use it as it makes sense
- Requires resources – time, personnel availability
  - Recommend 5+ days for detailed analysis
  - Invite the test team, operators, system SMEs
  - Bring in STPA experts if possible
  - In-person participation highly recommended
Questions?