How to Build Synthetic Persons in Cyberspace

Fernando Maymí
Alex Nickels

SOARTECH
Modeling human reasoning.
Enhancing human performance.

About Us

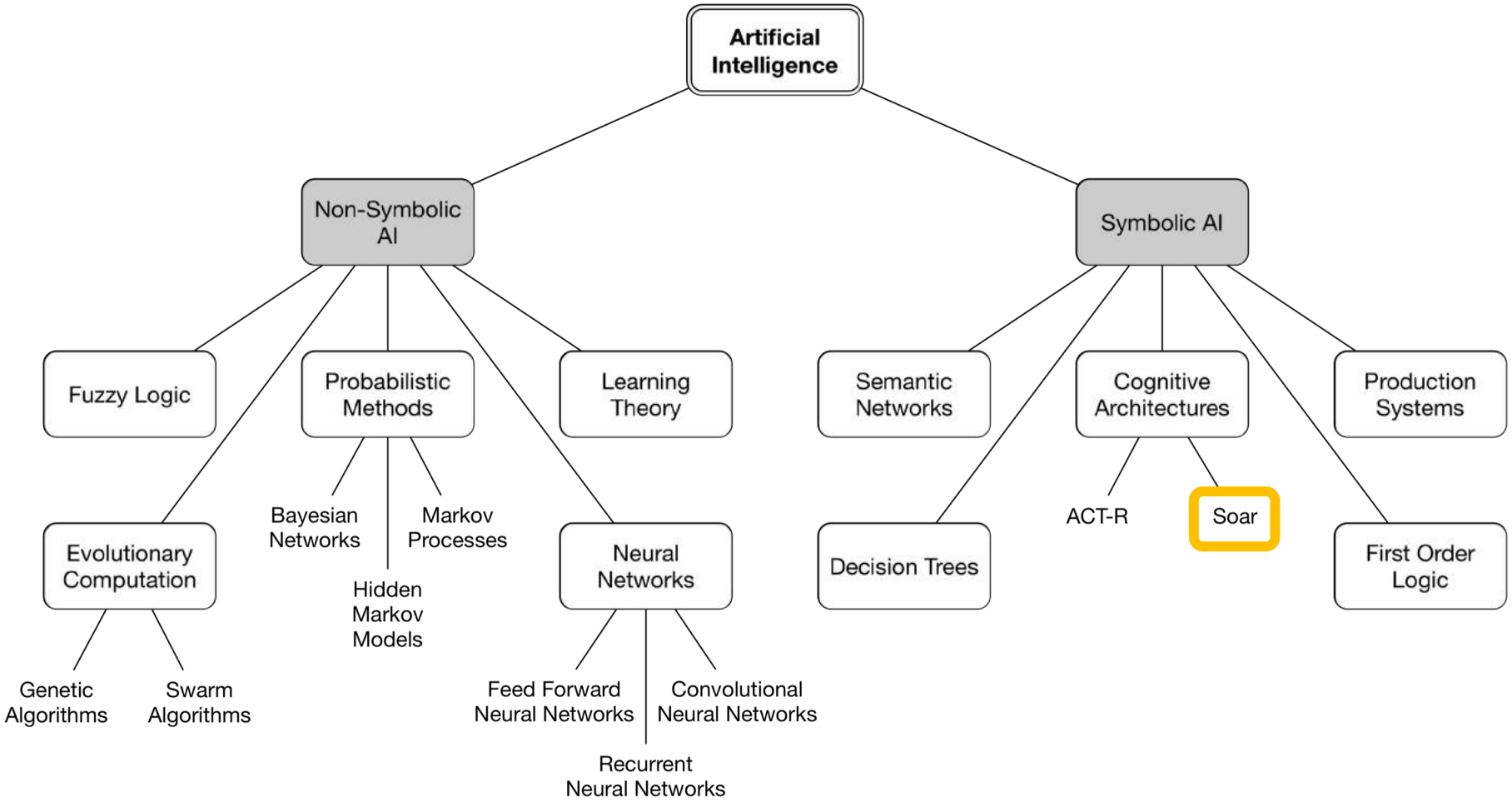
Fernando Maymí
Lead Scientist

Alex Nickels
Associate Technical Director

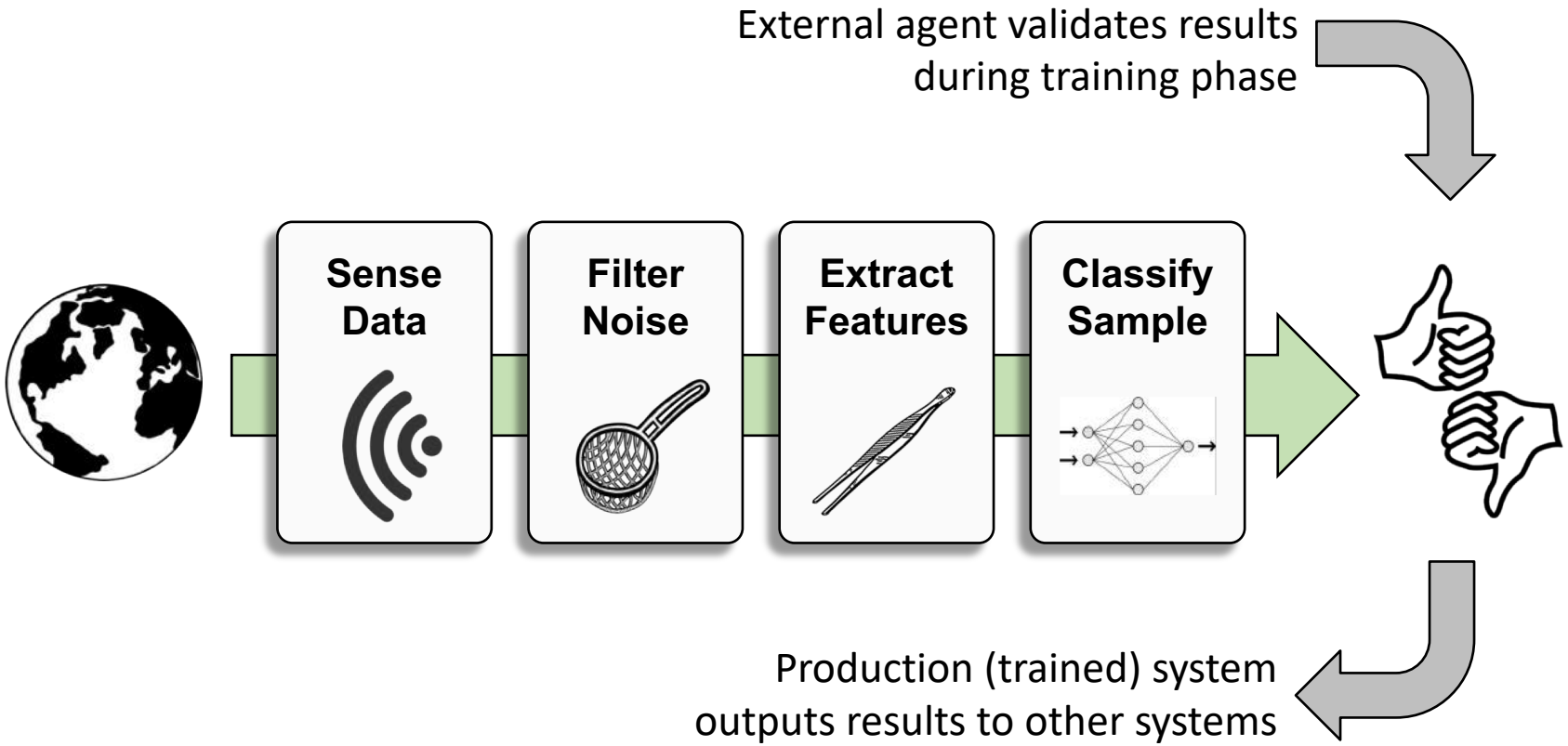
Agenda

- (Brief) Overview of Artificial Intelligence
- AI-related threats
- Cyberspace Cognitive (CyCog) agents
  - Attacker
  - Defender
  - User
- Future work

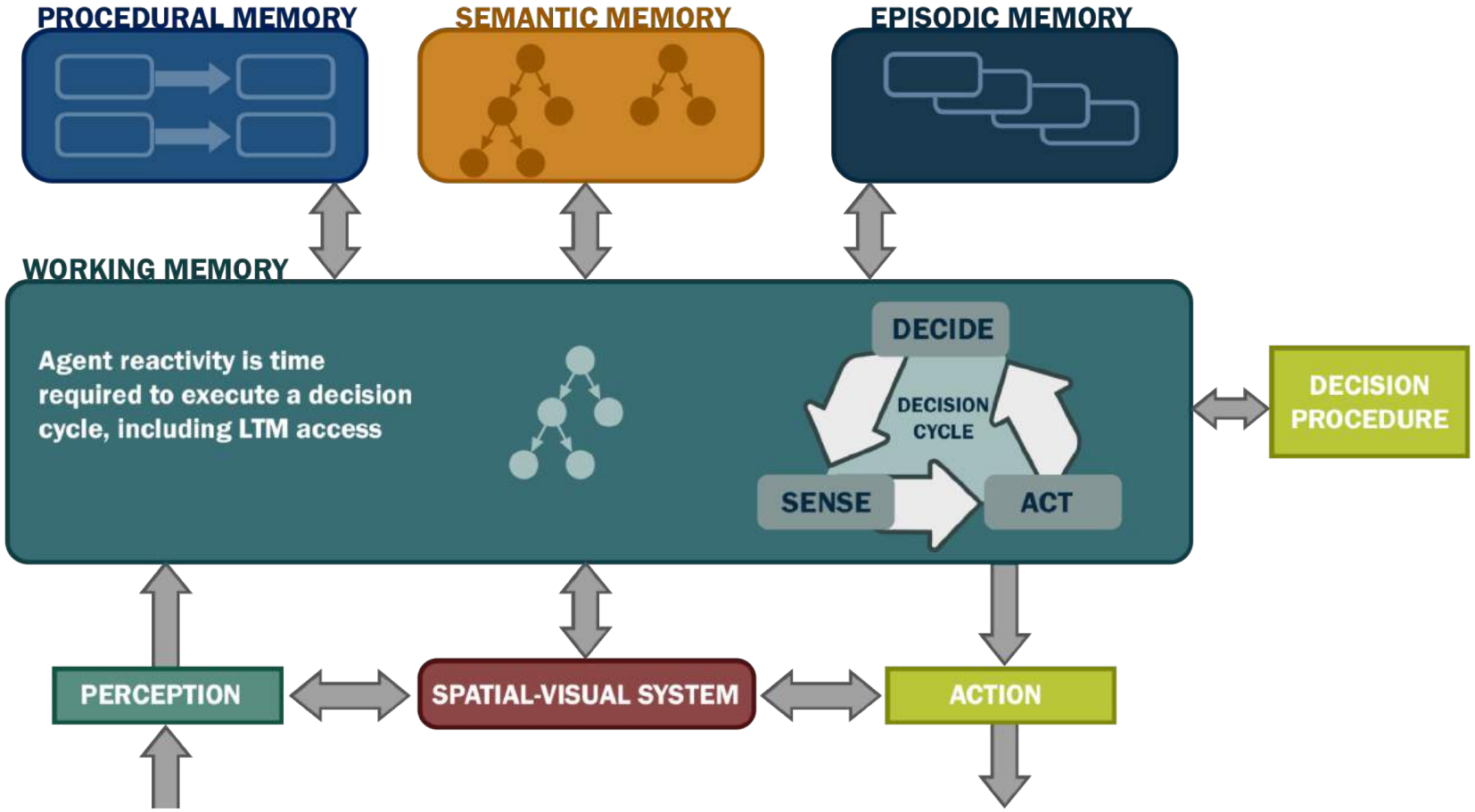
(Partial) AI Taxonomy



Non-Symbolic AI Example



Symbolic AI Example

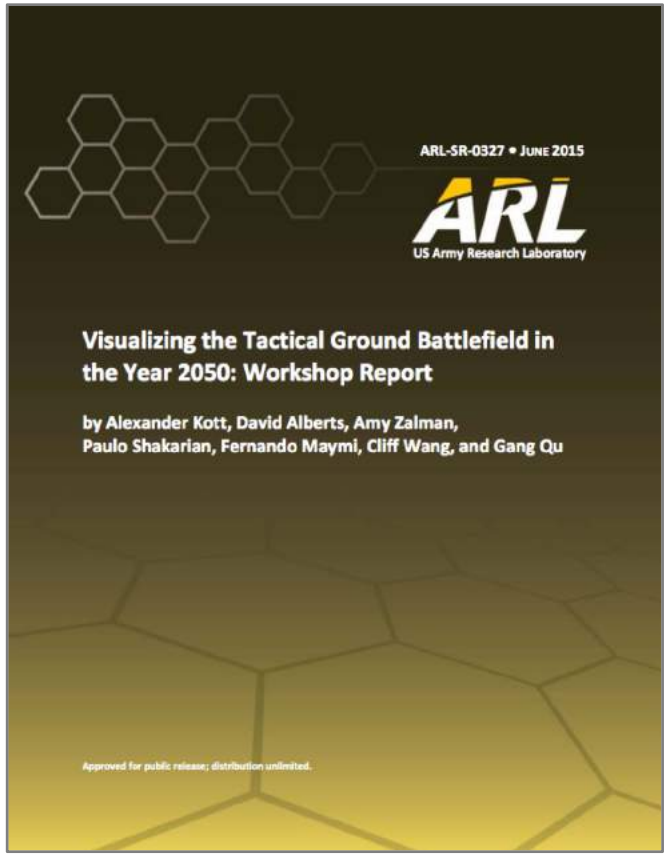


The open-source Soar Cognitive Architecture

AI-related Threats

The Tactical Battlefield of 2050

- Ability to understand and cope in a contested, imperfect, information environment
- Augmented humans
- Misinformation as a weapon
- Micro-targeting
- Large-scale self-organization and collective decision making
- Automated decision making and autonomous processes
- Cognitive modeling of the opponent


ARL-SR-0327 • June 2015

ARL
US Army Research Laboratory

Visualizing the Tactical Ground Battlefield in the Year 2050: Workshop Report

by Alexander Kott, David Alberts, Amy Zalman, Paulo Shakarian, Fernando Maymi, Cliff Wang, and Gang Qu

Approved for public release; distribution unlimited.

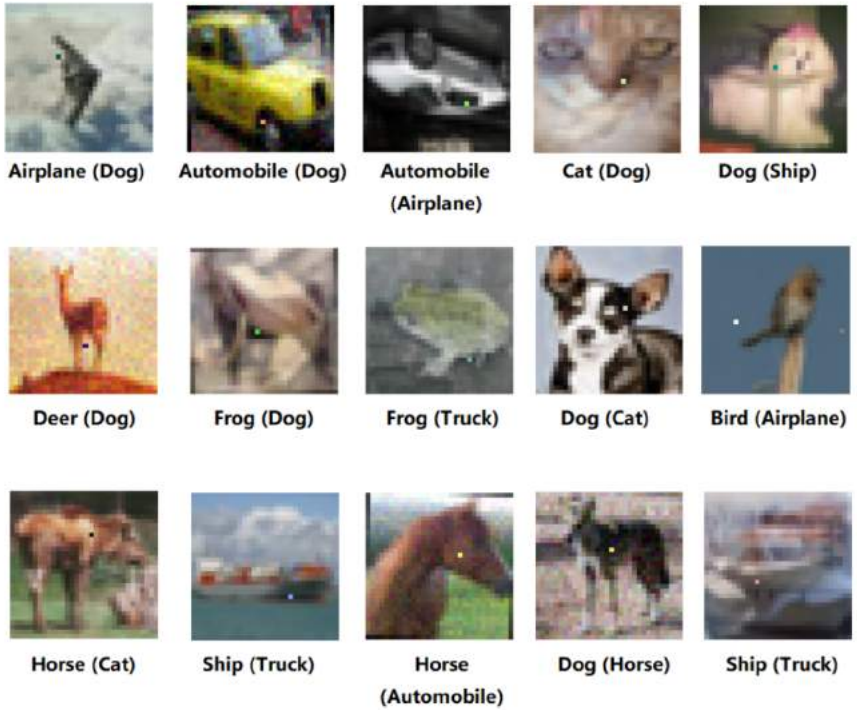
Imperfect AI

Adversarial Training

Microsoft's Tay

source: en.wikipedia.org/wiki/Tay_(bot)

Adversarial Machine Learning



One pixel attack for fooling deep neural networks

source: arxiv.org/pdf/1710.08864v2.pdf

Unintended Bias

Google results for "three black teens" v. "three white teens"

source: www.theguardian.com/commentisfree/2016/jun/10/three-black-teenagers-google-racist-tweet

AI and Cyber Attacks

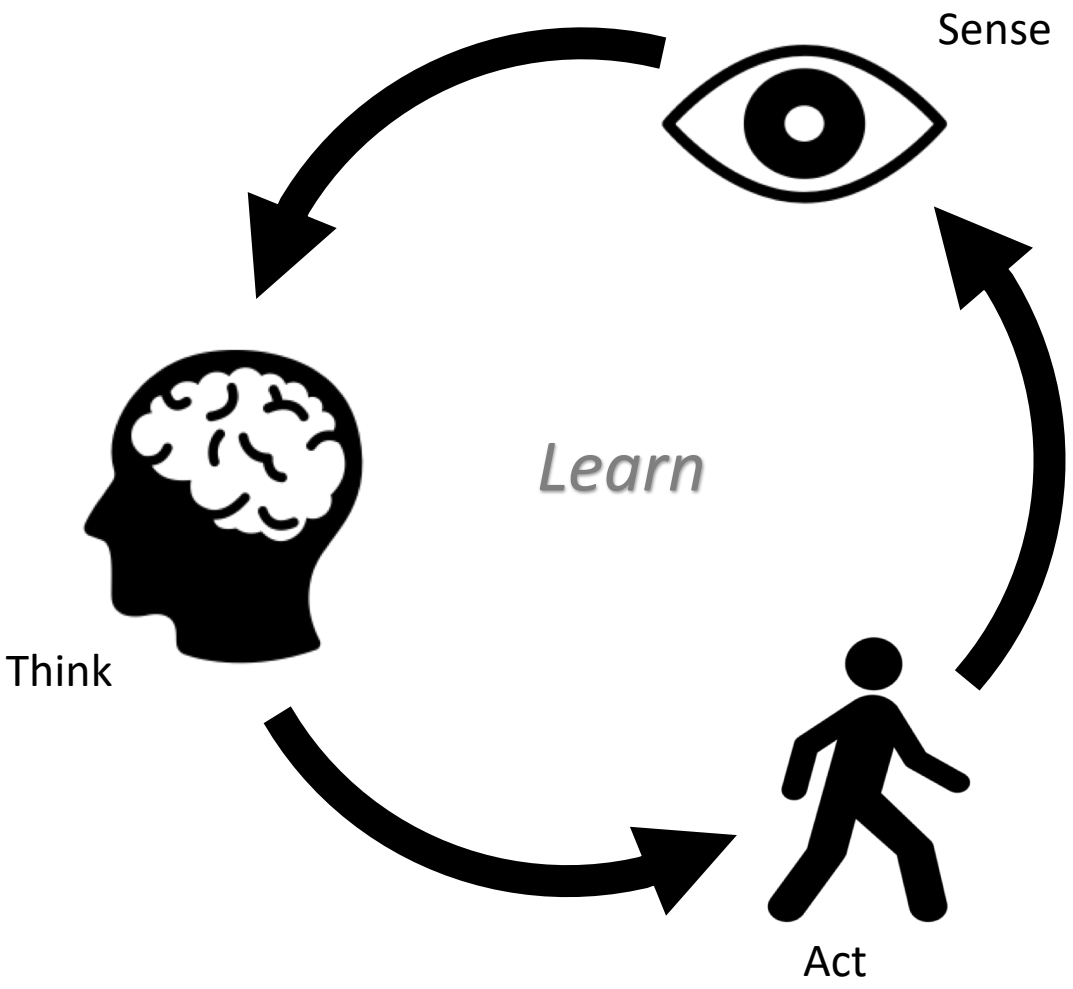
- IBM's Deep Locker (2018)
- DeepExploit (2018)
- Darktrace-reported attack in India (2017)
- DARPA Cyber Grand Challenge (2016)
- SNAP_R (2016)
- Death by Captcha
- Sentry MBA

DARPA Cyber Grand Challenge

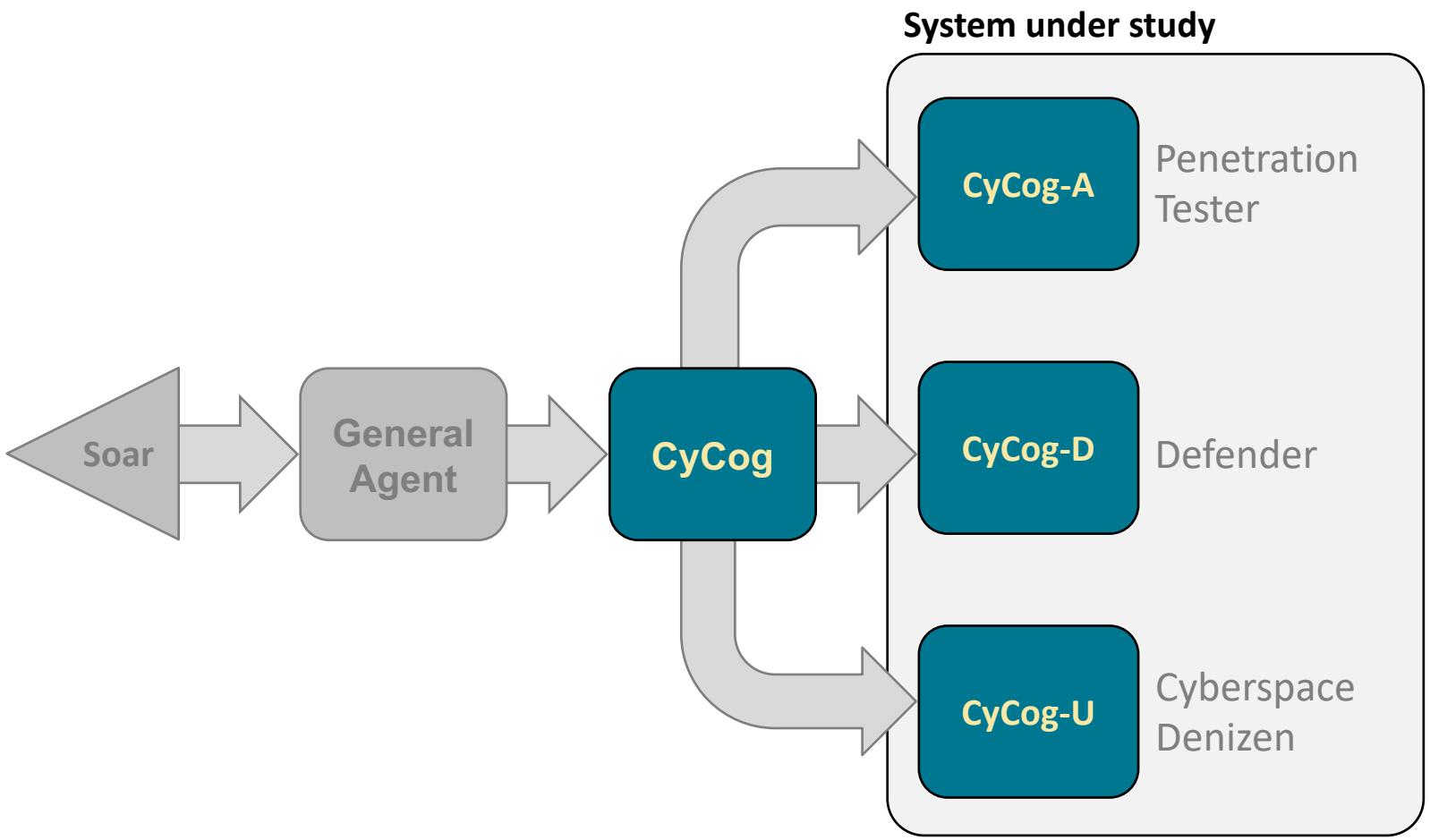
Source: www.darpa.mil

CyCog Agents

Autonomous Agents



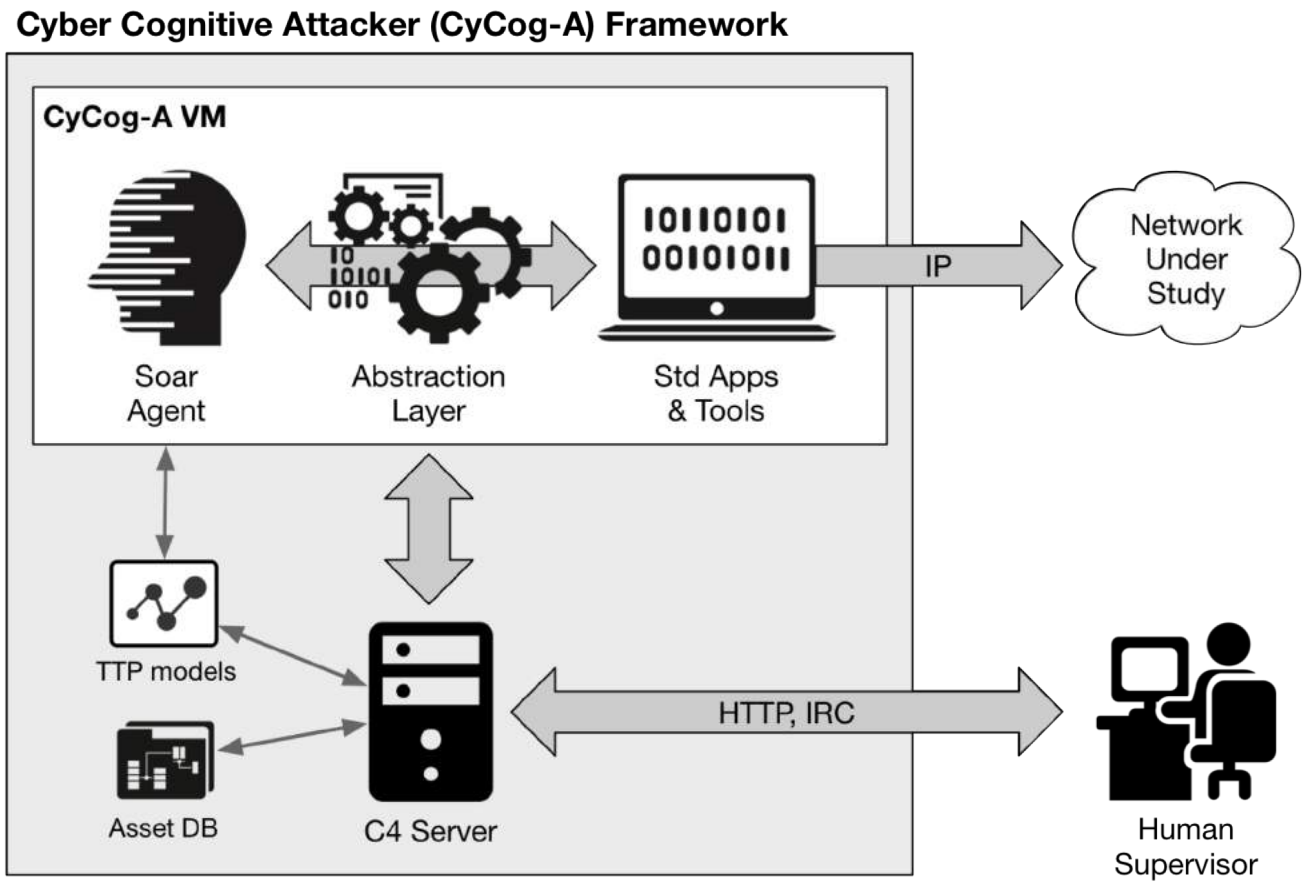
CyCog Agent Genealogy



CyCog: Cyber Cognitive
TTP: Tactics, Techniques & Procedures

CyCog-A Concept

- Soar agent models real-world adversary TTPs and interacts with standard tools to display realistic attacker behaviors against a network under study
- Currently able to use phishing, SQLi and remote exploitation to establish foothold, and then persists, moves laterally, searches & exfiltrates files



CyCog-A User Interface

SC2RAM

Plan

Execute

Review

Config

Running: Agent is "attacker-agent". MissionId is "attacker-agent".

server: attacker-agent start

Submit

Tree Console Debug Network Viz

Possess Information
Possess Information on System
Maintain Access to System
Possess Information Using System Access
Select System Access Method
Achieve Access To System
Use User Credential Access Method
Possess Credentials For System
Identify Vulnerable Host
Possess Credentials For Host
Identify Target Host By Service Type
Select Target Service Type
Identify Candidate Hosts
Find Credentials in Information
Possess Information
Identify Target Host Http Service
network-scan
Find Credentials In SQL Table Information
Possess Information on System
Identify Candidate URLs
Identify Vulnerable URLs
Possess Target Table Rows
Identify Target Table Fields
Possess Information Using System Access
Maintain Access to System

Tree Console Debug

```
[(no context provided)][15:46:41] [INFO] resumed: t
[(no context provided)][15:46:41] [INFO] resumed: 3
[(no context provided)][15:46:41] [INFO] resumed: 0
[(no context provided)][15:46:41] [INFO] resumed: 0
[(no context provided)][15:46:41] [INFO] resumed: batman
[(no context provided)][15:46:41] [INFO] resumed: batman
[(no context provided)]
[(no context provided)]do you want to store hashes to a temporary file for eventual further processing with other tools [y/N] N
[(no context provided)]
[(no context provided)]do you want to crack them via a dictionary-based attack? [Y/n/q] N
[(no context provided)]Database: evilforum
[(no context provided)]Table: evil_users
[(no context provided)][3 entries]
[(no context provided)]+---------+----------+-------------------+-------------------+--------------+----------+----------+----------+----------+----------+----------+-------
[(no context provided)]| user_id | group_id | user_avatar_width | user_sig_bbcode_uid | user_ip      | user_new | user_sig | username | user_lang | user_rank | user_type | user_p
[(no context provided)]+---------+----------+-------------------+-------------------+--------------+----------+----------+----------+----------+----------+----------+-------
[(no context provided)]| 1       | 1        | 0                 | <blank>           | <blank>      | 1        | <blank>  | Anonymous | en       | 0        | 2        | 0
[(no context provided)]| 2       | 5        | 0                 | <blank>           | 10.0.0.6     | 1        | <blank>  | admin    | en       | 1        | 3        | 1
[(no context provided)]| 2348    | 5        | 0                 | <blank>           | 192.168.1.10 | 1        | <blank>  | batman   | en       | 1        | 3        | 1
[(no context provided)]+---------+----------+-------------------+-------------------+--------------+----------+----------+----------+----------+----------+----------+-------
[(no context provided)]
```

CyCog-D User Interface

CYCOG
Defender

Execute

Review

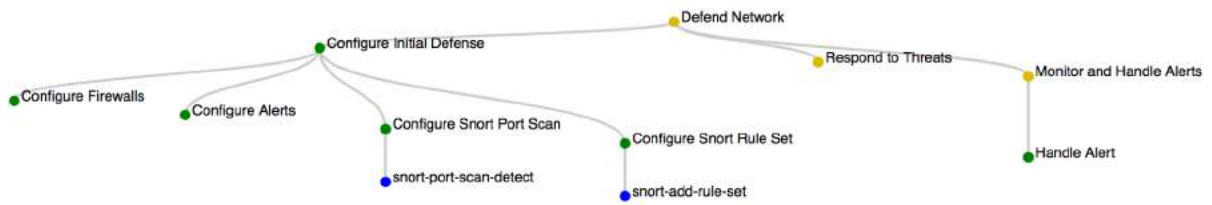
Config

Agent no longer connected: Agent was "defender-agent". MissionId was "defender-agent".

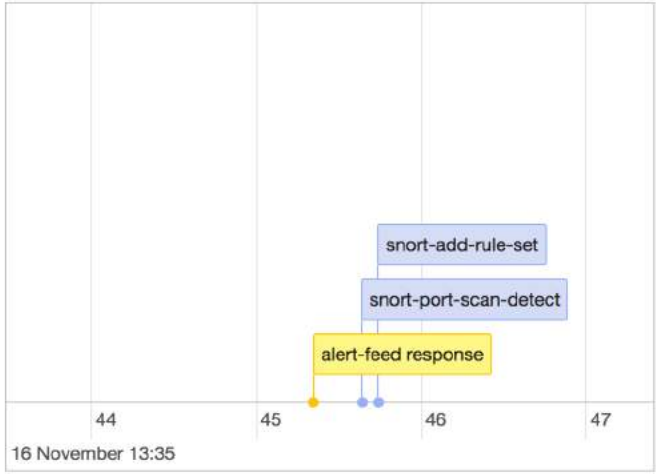
cycog-agent: cycog-agent connected

defender-agent: defender-agent connected

Submit



Timeline Console Debug

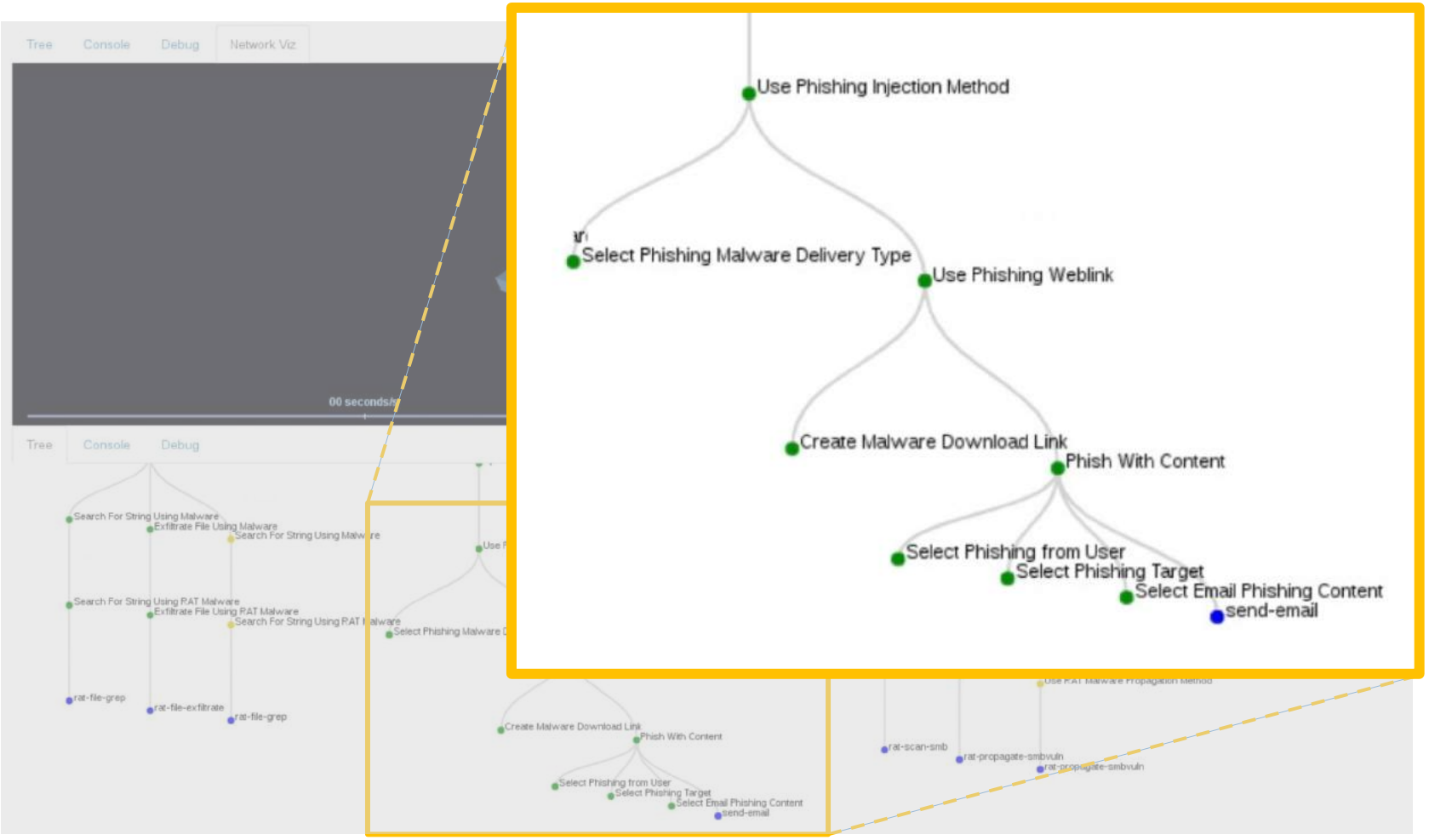


alert-feed response

Triggered by: Unknown


```
{
  "source": {
    "netmask": {
      "inetAddress": "255.255.255.0",
      "four": 0,
      "one": 255,
      "three": 255,
      "two": 255
    },
    "ip": {
      "inetAddress": "192.168.57.0",
      "four": 0,
      "one": 192,
      "three": 57,
      "two": 168
    }
  },
  "id": "00c87b6d-ff51-4ea5-8942-ee0a5db765ec",
  "message": "Malicious traffic originating from subnet",
  "type": "threat-feed",
  "timestamp": 1510857344273
}
```

Inherently Explainable AI



Modeling TTPs

Modeling TTPs

RaGE | Goal Model Editor | Type Hierarchy Editor | Choose File knowledge.xml | Download As XML

Edit

Identify Candidate URLs

achieve

Identify Vulnerable URLs

achieve

Identify Target Host Http Service

achieve

Identify Target Host

Identify Target Host Http Service

Name

A unique identifier for this goal node that can be used to reference it in goal bindings/edges

identify-target-host-http-service

Description

A long-form description of the purpose of this goal

Identify the target host using HTTP service.

Achieved When

Define the conditions for when this goal has succeeded

NOTE: Rule groups and ORs are not currently supported by the goal model

AND | + Add rule | Add group

host-id | ✖ Delete

has value

Failed When

Define the conditions for when this goal has failed

NOTE: Rule groups and ORs are not currently supported by the goal model

Agent Code

RIDL

- XML Goal Declarations – Declarative language and interpreter for task goal hierarchies
- Soar Rules – Interpreter to execute goals

CyCog

- Soar Rules – Instructions to interact with the abstraction layer and data stores

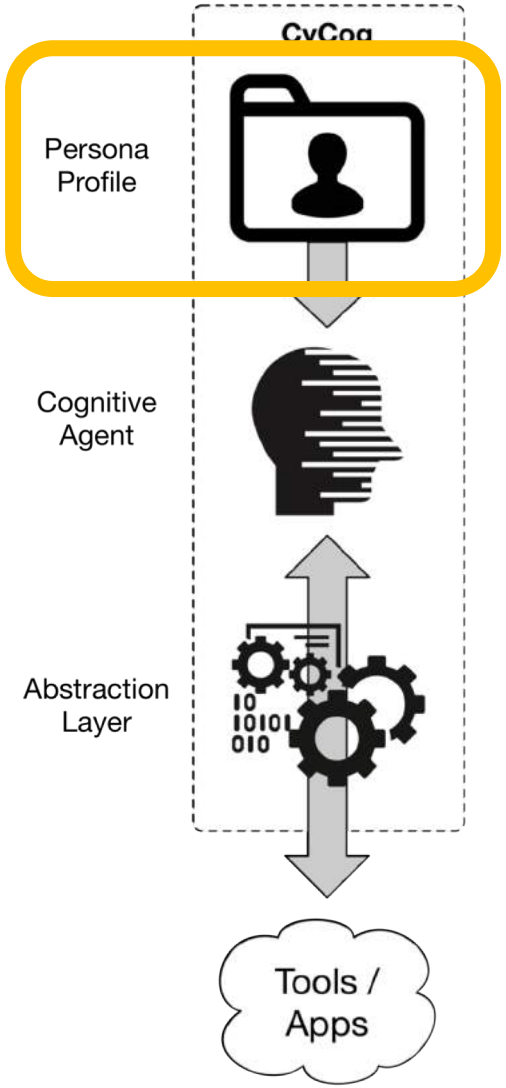
XML Goal Declarations

```
<goal>
    <name value="possess-information-using-network-access" />
    <metadata>
      <display-name value="Possess Information Using Network Access" />
      <description value="Attempt to acquire some information.  Acquire access to a host network as the means of accomplishing this." />
      <editor-x>-229</editor-x>
      <editor-y>-298</editor-y>
    </metadata>
    <parameter>
      <name value="host-id" />
    </parameter>
    <parameter>
      <name value="information-type" />
    </parameter>
    <parameter>
      <name value="information-locator-i
    </parameter>
    <parameter>
      <name value="information-locator-t
    </parameter>
    <parameter>
      <name value="information-id" />
    </parameter>
    <parameter>
      <name value="network-access-type
    </parameter>
    <parameter>
      <name value="network-access-achie
    </parameter>
    <achieved-when>
      <has-value>
        <parameter value="information-id
      </has-value>
    </achieved-when>
  </goal>
```

Soar Productions

```
sp {execute-command*apply*network-scan*result
    :o-support
    (state <s> ^operator <o> ^io <io>)
    (<io> ^input-link.result <ir> ^output-link <ol>)
    (<ol> ^network-scan <c>)
    (<o> ^name execute-command ^command-name network-scan ^command-id <id>
         ^result-name <rn> ^result-destination <rd>
         ## Copy only when results are complete
         ^okay-to-copy-result true)
    (<c> ^command-id <id>)
    (<ir> ^command-id <id>)
-->
    (write (crlf) |Got result for | <id> | from I/O system, storing to |
<rd> | ^| <rn>)
    (<rd> ^<rn> (deep-copy <ir>))
    (<ol> ^<cmd> <c> -)
}
```

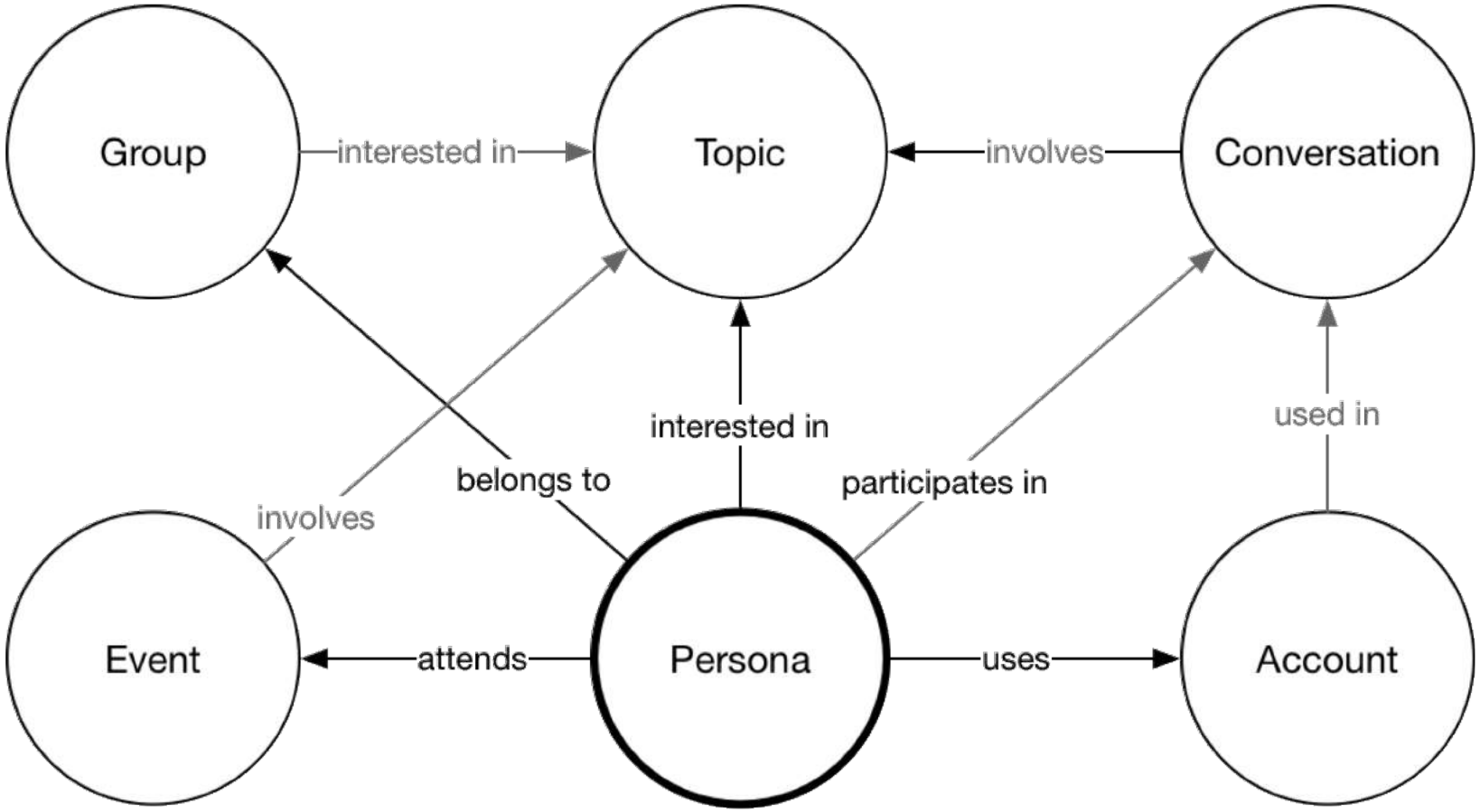
CyCog Personas



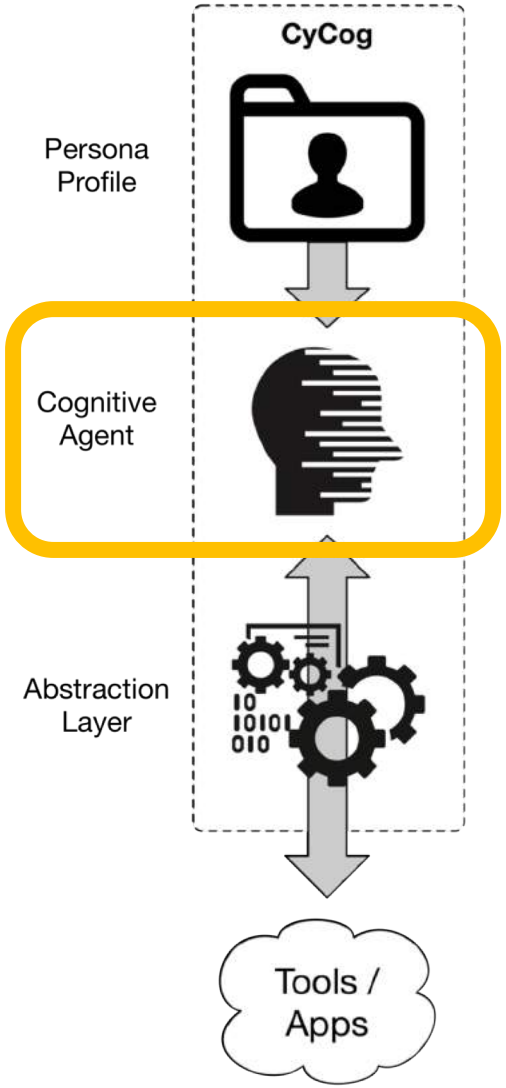
Persona Profiles

- Procedural knowledge
  - TTPs for CyCog-A and D
  - General activities (email, web browsing) for all
- Personal information
  - Demographics
  - Credentials
- Social affinities
  - Indicator of like/dislike for others
  - Could eventually model richer interactions
- Topical affinities
  - Personal interests
  - Emotional triggers

CyCog Personas



CyCog Agents



ACT-R

- Used for CyCog-U agents
- Use instance-based learning for adaptive behavior

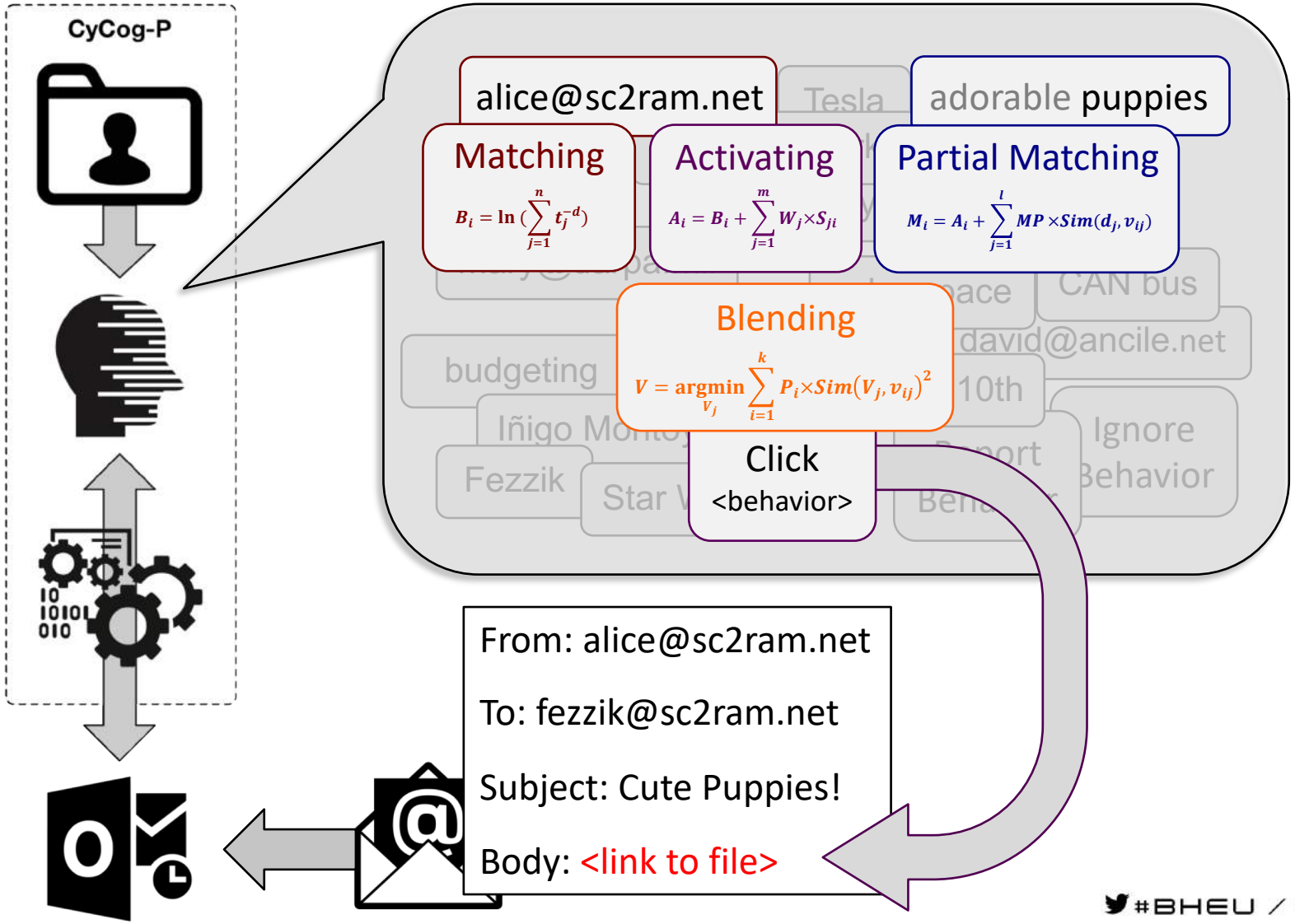
Soar

- Used by CyCog-A and D
- Episodic memory provides weak learning

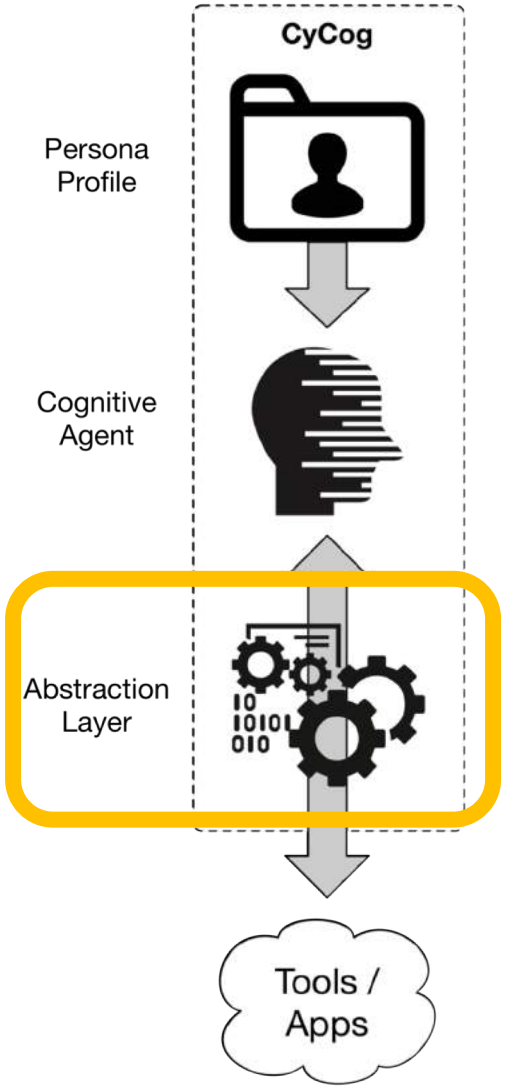
Java

- Lightweight CyCog-U agents for lo-fi scalability
- No learning

ACT-R Agent Learning



CyCog Abstraction Layer



Hacking Toolkit

- Demonstrated with multiple hacking tools
- Building tool interfaces requires significant work

Defensive Toolkit

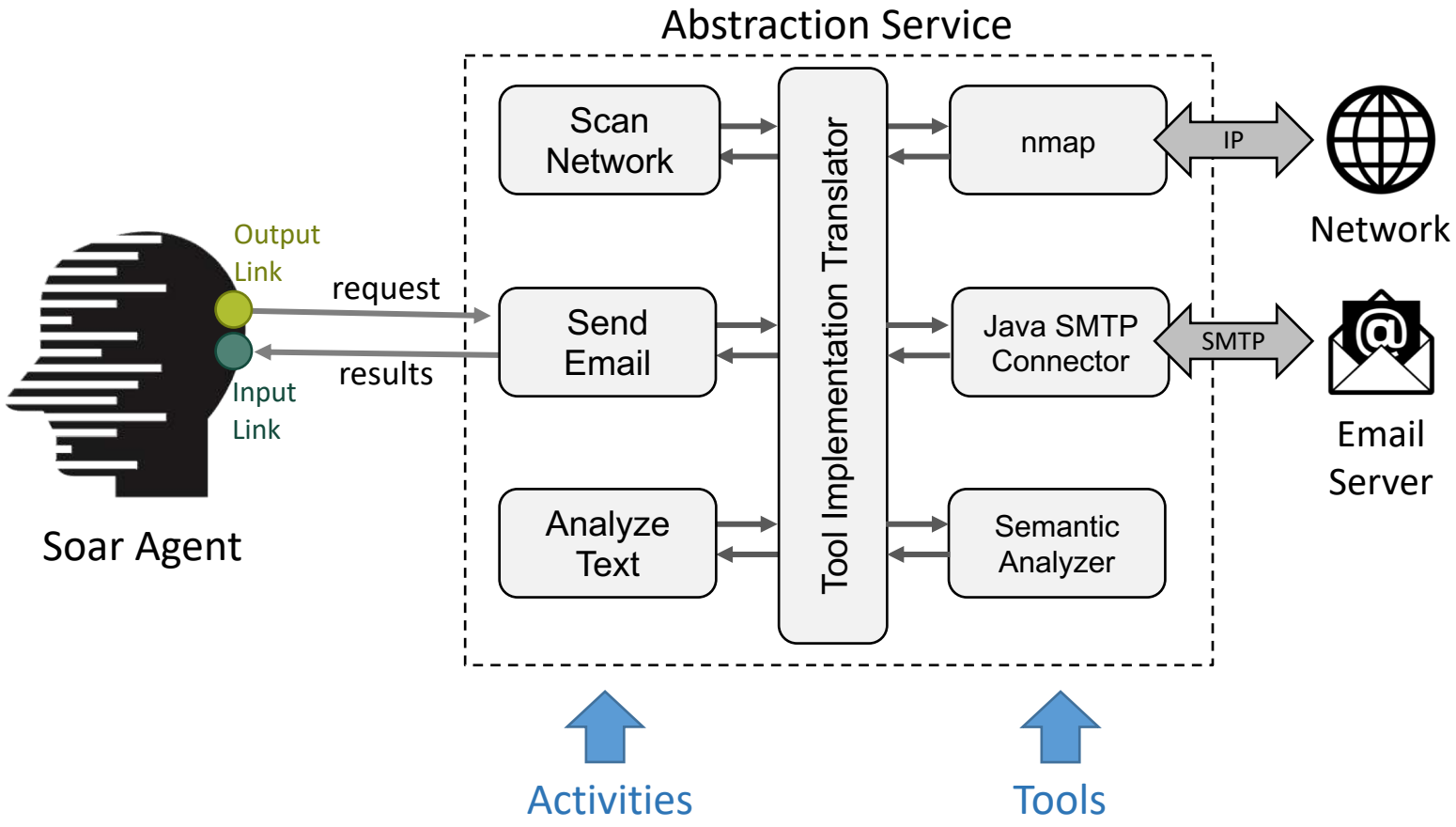
- Uses OpenC2
- Knows how to use iptables, Snort

CyCog-U

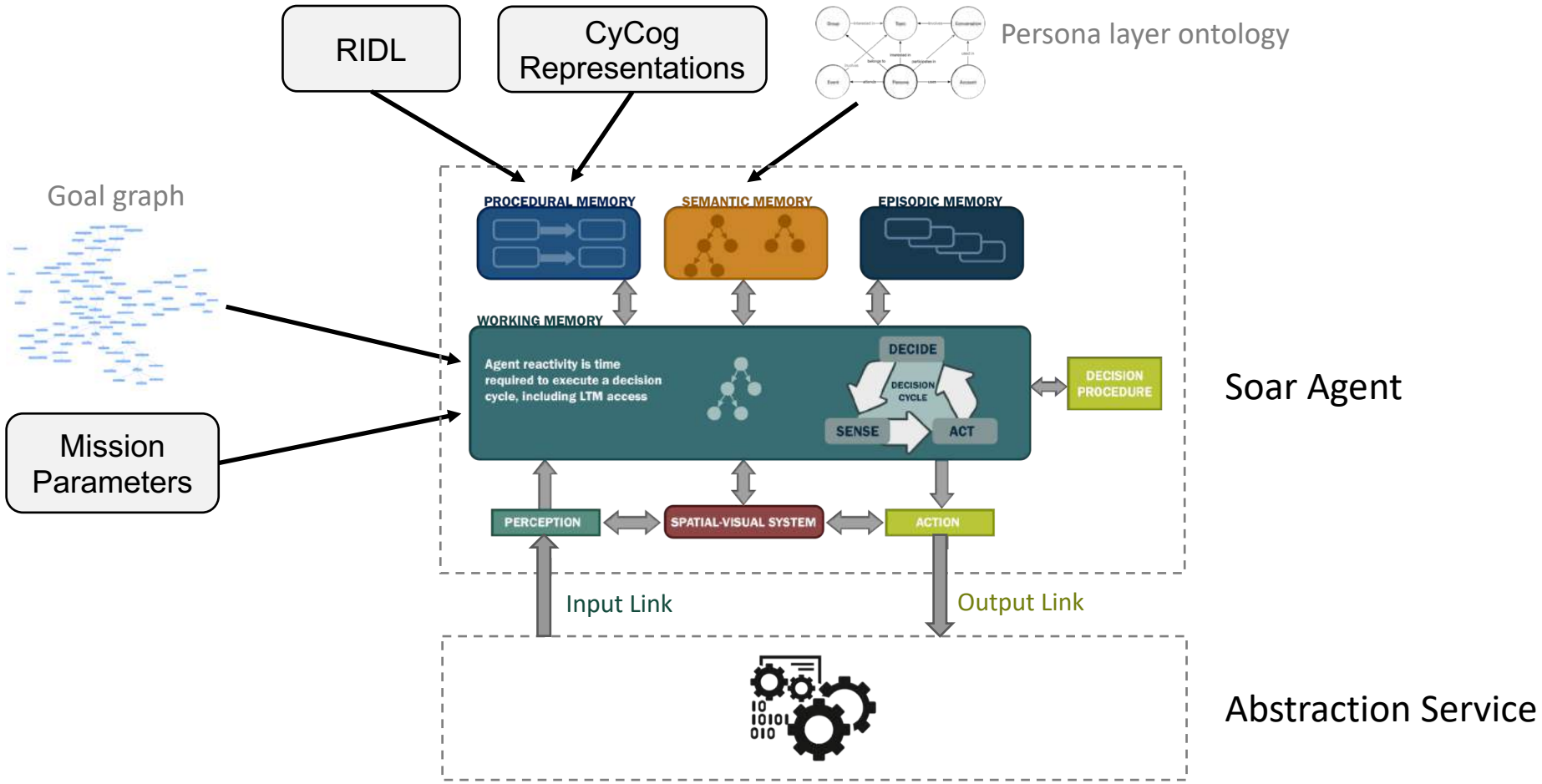
- Very basic natural language processing
- Email reading
- Web browsing

CyCog Abstraction Layer

1. Agent requests an activity be performed to send email using email server
2. Abstraction service finds tool able to perform the activity and runs the tool
3. Implementation translator turns results into format matching original activity
4. Abstraction service returns results to the agent



Putting It All Together



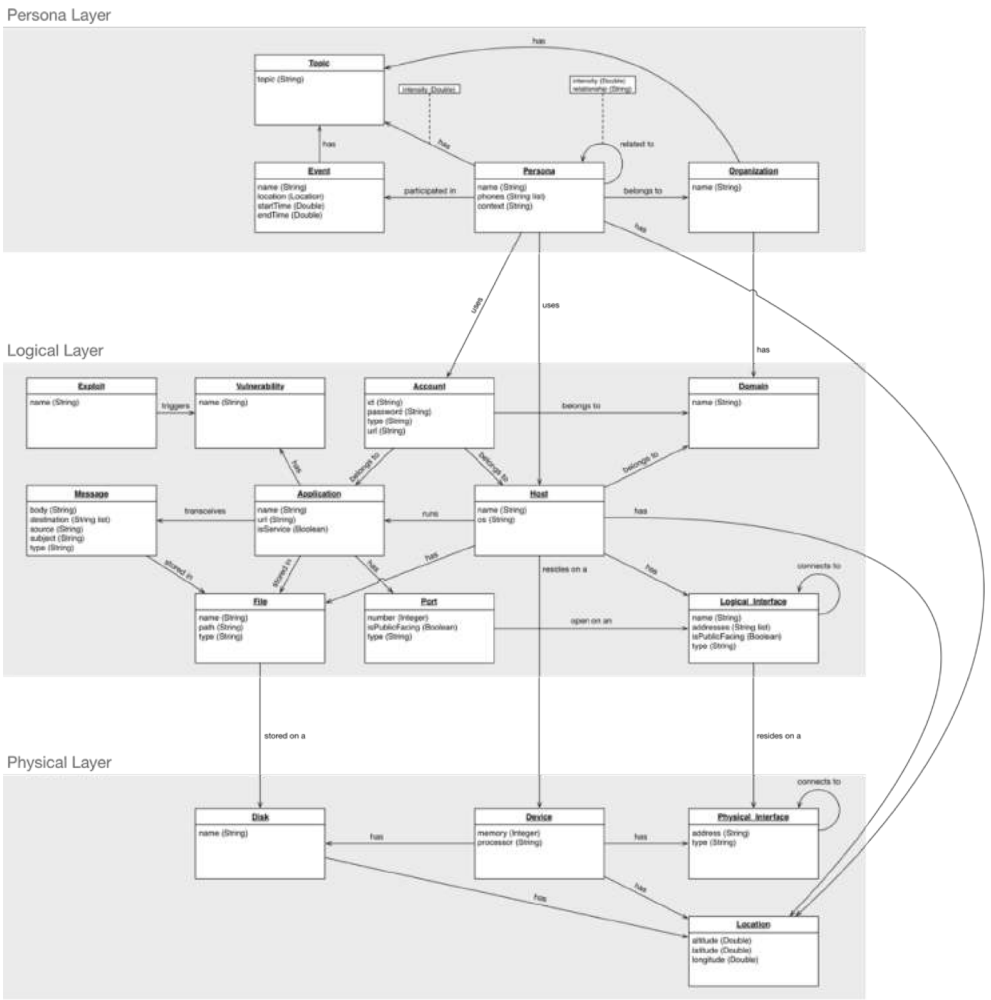
Demo

~~Future~~ Work

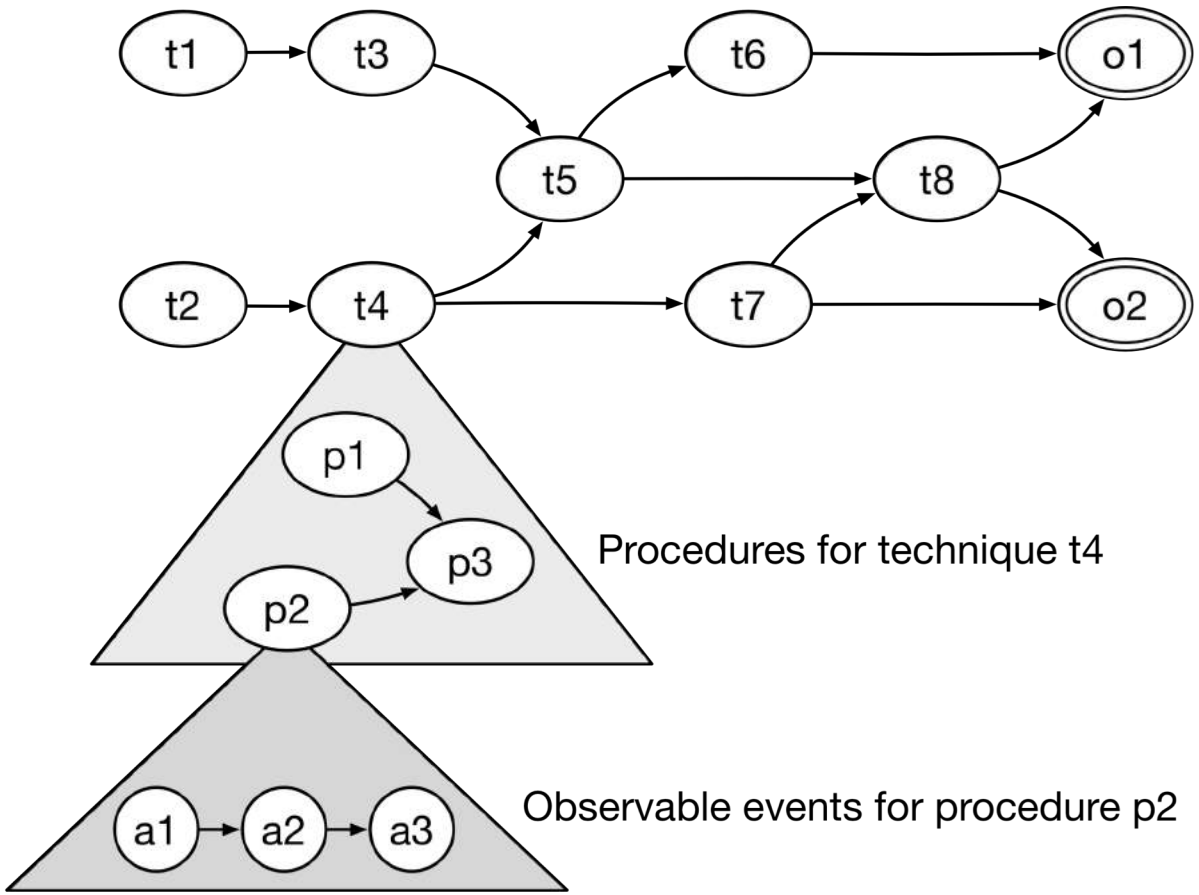
Keeping track of cyberstuff
Modeling TTPs
Human-machine teaming

Keeping Track of Cyberstuff

Cyberspace Layer	Modeling Aspects
Cyber-Persona (Cognitive/ Social) 	• Personas and Identities (many-to-many) • Intent/Goals • TTPs, C2 • Social presence and communication
Logical 	• Operating system + drivers • Applications (to include malware) • Network protocols • Events and Logs
Physical 	• Hardware architecture • Physical compute nodes • Physical network connections • Geo-Location of compute nodes • Persona biometrics (key stroke, mouse patterns, facial recognition)



Formal TTP Models



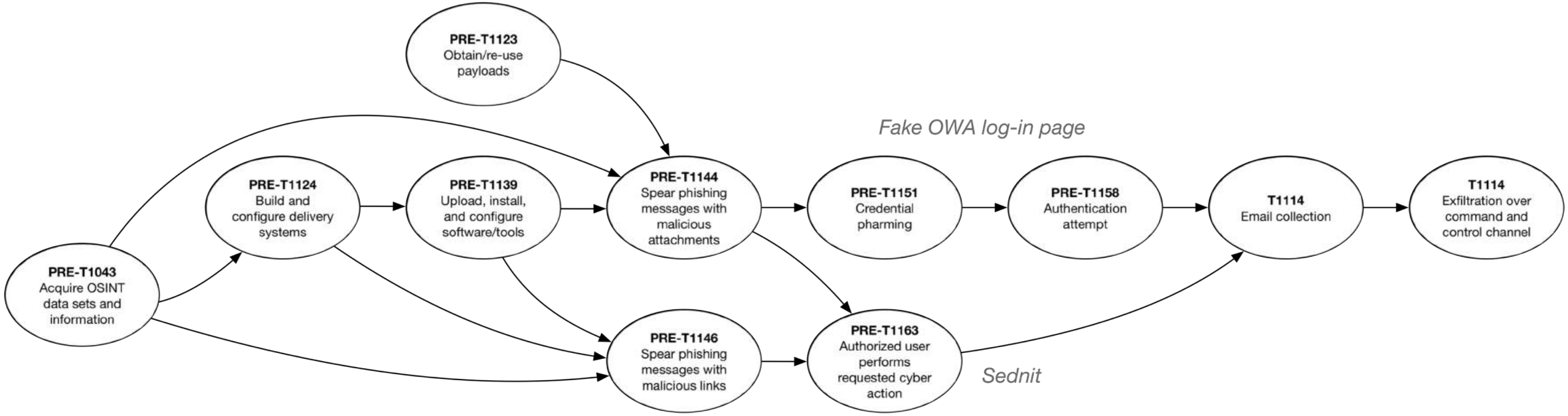
Tactic –employment and ordered arrangement of forces in relation to each other

Technique – non-prescriptive way to perform missions, functions, or tasks

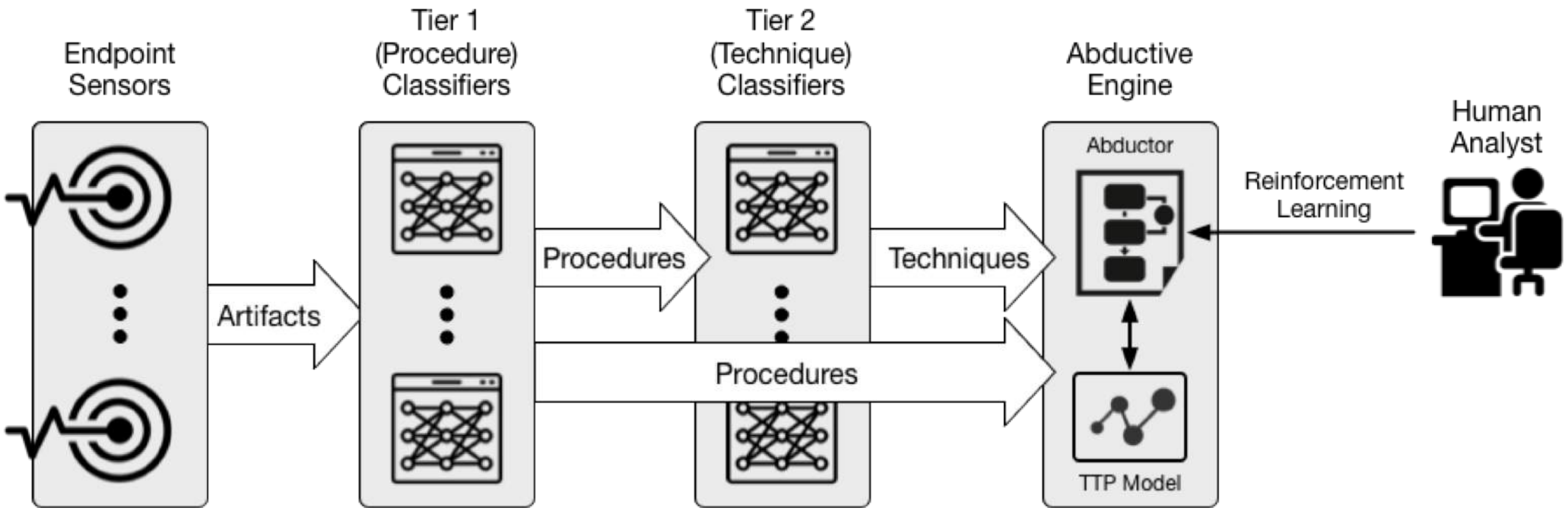
Procedure – standard, detailed steps that prescribe how to perform specific tasks

TTP Model + ATT&CK

Operation Pawn Storm (APT28)



Using ML to Build TTP Models



SOARTECH

Modeling human reasoning.
Enhancing human performance.

fernando.maymi@soartech.com
alex.nickels@soartech.com