Relational Reinforcement Learning for Automated Network Intrusion Response

Jakob Nyberg
May 21, 2026 — KTH Royal Institute of Technology

Sentience

- Funded by MCF.
- Project Goal: Semi-automated SOC.

Sentience

- Funded by MCF.
- Project Goal: Semi-automated SOC.

Today

- Test agents for automated cyber defense.

Task Formulation

- Hardening.
- Detection.
- **Mitigation**.
- Recovery.

Task Formulation

We formulate mitigation as a Markov decision process.

- Network hosts being accessed by an unauthorized actor costs us *something*.

Task Formulation

We formulate mitigation as a Markov decision process.

- Network hosts being accessed by an unauthorized actor costs us *something*.
- Find actions to prevent it.

Task Formulation

We formulate mitigation as a Markov decision process.

- Network hosts being accessed by an unauthorized actor costs us *something*.
- Find actions to prevent it.
- Actions costs us *something*.

Task Formulation

We formulate mitigation as a Markov decision process.

- Network hosts being accessed by an unauthorized actor costs us *something*.
- Find actions to prevent it.
- Actions costs us *something*.
- Agent should minimize the joint cost.

Agents

- Relational RL agent(s) trained using attack simulation.
- Heuristic agent.
- Do nothing.

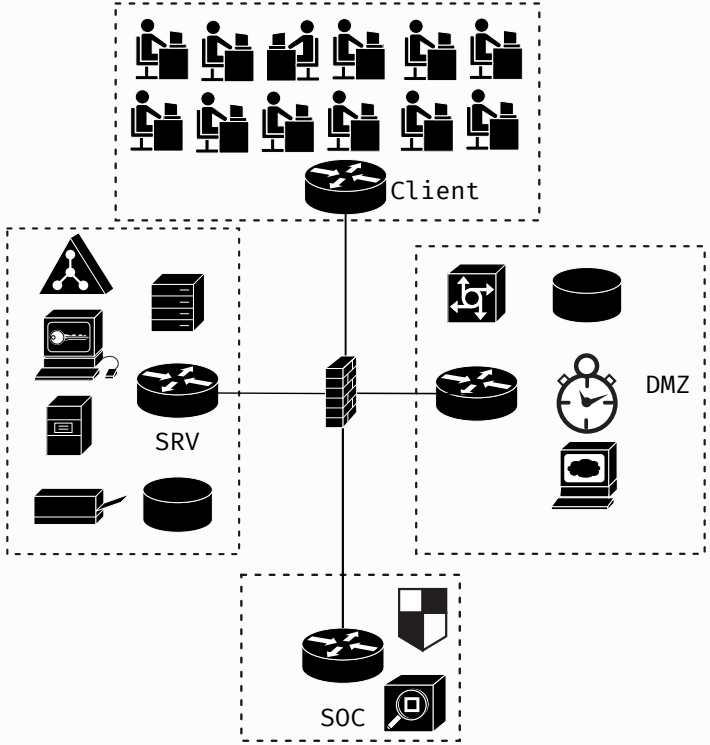
CRATE

- Developed and maintained by FOI.
- Cyber range used for cyber security exercises.

CRATE

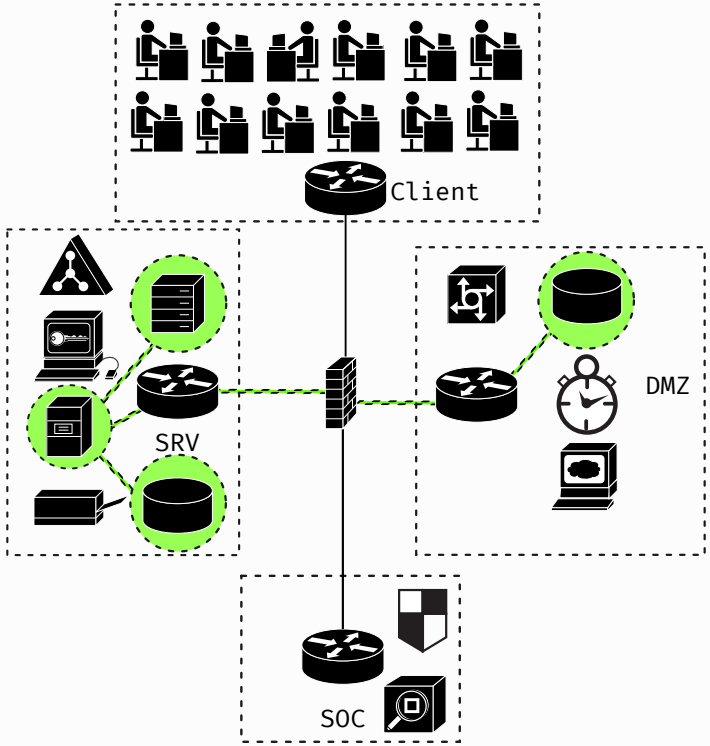
- Developed and maintained by FOI.
- Cyber range used for cyber security exercises.
- Red-team emulation using Lore.
- Simulated user events.

AIR-DELIVERY-SYSTEM24



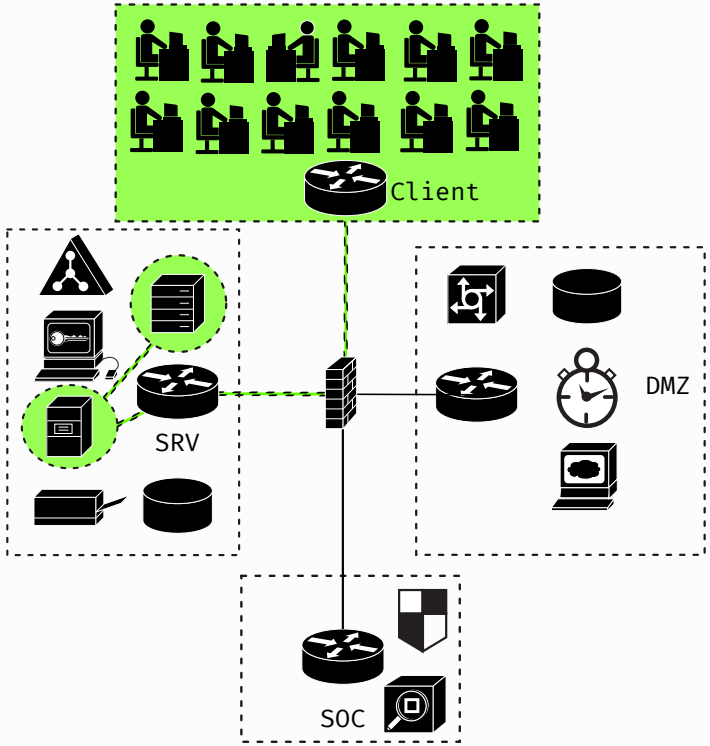
AIR-DELIVERY-SYSTEM24

Guided Strategy

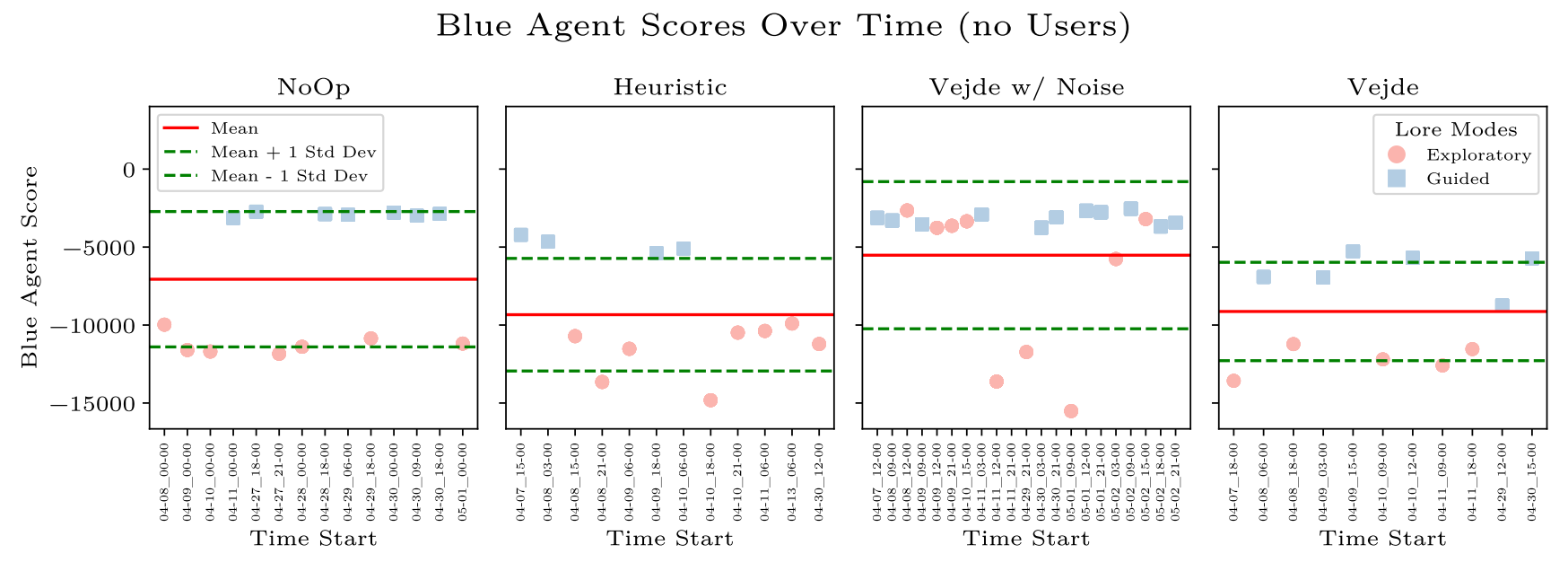


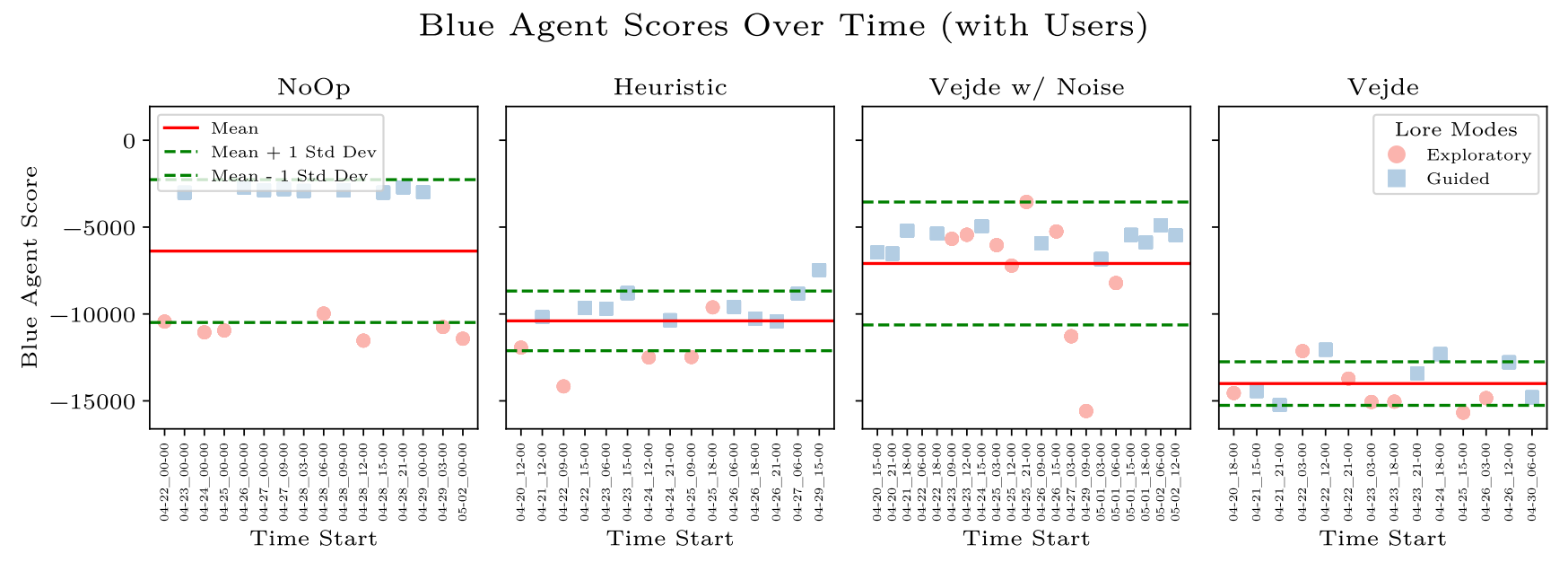
AIR-DELIVERY-SYSTEM24

Exploratory Strategy



Host Class	Availability	Integrity	Confidentiality
Time Server / NTP	4	5	1
Log Server	5	5	4
File Server	4	2	2
Domain Controller	5	5	2
Name Server	5	5	2
Web Server	2	4	1
CA-Server	5	5	1
Clients	5	5	2
Mail Server	3	4	3
Mail Relay	3	4	3
Payroll Server	5	5	2
DB Server	5	5	2


Blue Agent Scores Over Time (no Users)
NoOp
Heuristic
Vejde w/ Noise
Vejde
Mean
Mean + 1 Std Dev
Mean - 1 Std Dev
Lore Modes
Exploratory
Guided
Blue Agent Score
0
−5000
−10000
−15000
Time Start
04-08_00-00
04-09_00-00
04-10_00-00
04-11_00-00
04-27_18-00
04-27_21-00
04-28_00-00
04-28_18-00
04-29_06-00
04-29_18-00
04-30_00-00
04-30_09-00
04-30_18-00
05-01_00-00
04-07_15-00
04-08_03-00
04-08_15-00
04-08_21-00
04-09_06-00
04-09_18-00
04-10_06-00
04-10_18-00
04-10_21-00
04-11_06-00
04-13_06-00
04-30_12-00
04-07_12-00
04-08_09-00
04-08_12-00
04-09_09-00
04-09_12-00
04-09_21-00
04-10_15-00
04-11_03-00
04-11_12-00
04-11_21-00
04-29_21-00
04-30_03-00
04-30_21-00
05-01_09-00
05-01_12-00
05-01_21-00
05-02_03-00
05-02_09-00
05-02_15-00
05-02_18-00
05-02_21-00
04-07_18-00
04-08_06-00
04-08_18-00
04-09_03-00
04-09_15-00
04-10_09-00
04-10_12-00
04-11_09-00
04-11_18-00
04-29_12-00
04-30_15-00


Blue Agent Scores Over Time (with Users)
NoOp
Heuristic
Vejde w/ Noise
Vejde
Blue Agent Score
0
−5000
−10000
−15000
Mean
Mean + 1 Std Dev
Mean - 1 Std Dev
Lore Modes
Exploratory
Guided
Time Start
04-22_00-00
04-23_00-00
04-24_00-00
04-25_00-00
04-26_00-00
04-27_00-00
04-27_09-00
04-28_03-00
04-28_06-00
04-28_09-00
04-28_12-00
04-28_15-00
04-28_21-00
04-29_00-00
04-29_03-00
05-02_00-00
04-20_12-00
04-21_12-00
04-22_09-00
04-22_15-00
04-23_06-00
04-23_15-00
04-24_12-00
04-24_21-00
04-25_09-00
04-25_18-00
04-26_06-00
04-26_18-00
04-26_21-00
04-27_06-00
04-29_15-00
04-20_15-00
04-20_21-00
04-21_18-00
04-22_06-00
04-22_18-00
04-23_09-00
04-23_12-00
04-24_15-00
04-25_03-00
04-25_12-00
04-25_21-00
04-26_09-00
04-26_15-00
04-27_03-00
04-29_09-00
05-01_03-00
05-01_06-00
05-01_15-00
05-01_18-00
05-02_06-00
05-02_12-00
04-20_18-00
04-21_15-00
04-21_21-00
04-22_03-00
04-22_12-00
04-22_21-00
04-23_03-00
04-23_18-00
04-23_21-00
04-24_18-00
04-25_15-00
04-26_03-00
04-26_12-00
04-30_06-00

Open Questions

- Security implications?
- Is a cost function too limited?
- Time scale?

But what about LLMs?

1. Feed Wazuh log into LLM.
2. Add instructions to act when needed.
3. ???
4. Success?

But what about LLMs?

1. Feed Wazuh log into LLM.
2. Add instructions to act when needed.
3. ???
4. Success?

Problem: Context Length

- 2 hours of Wazuh data in JSON $\approx 13\,332\,857$ tokens.

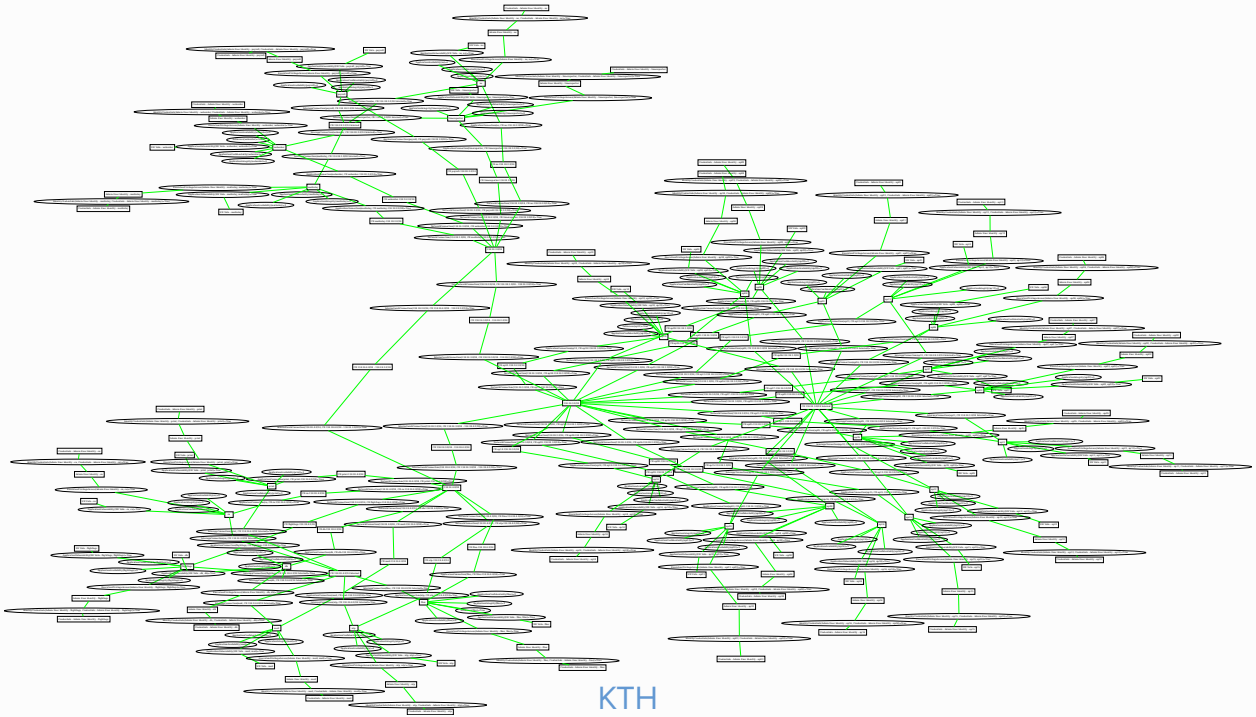
But what about LLMs?

1. Feed Wazuh log into LLM.
2. Add instructions to act when needed.
3. ???
4. Success?

Problem: Context Length

- 2 hours of Wazuh data in JSON $\approx 13\,332\,857$ tokens.
- Data modeling. MAL data model $\approx 311\,828$ tokens.

Knowledge Graph




KTH
VETENSKAP
OCH KONST