Enhancing Physical-based CPS Anomaly Detection Using Large Language Models

Shaikh Islam
sislam4@charlotte.edu

Chenglong Fu
chenglong.fu@charlotte.edu

Meera Sridhar
msridhar@charlotte.edu

Motivation

- Challenges of CPS Invariant Extraction

- Require large amount of data → Slow
- Require domain-specific expertise → Costly
- Diverse on different CPS → Hard to automate.

Core Insight

- Use of large language models (LLMs)
- Extract semantic information from CPS documents
- Generate hypothetical invariants from LLM
- Verify hypothetical invariants using a small amount of data

Contribution

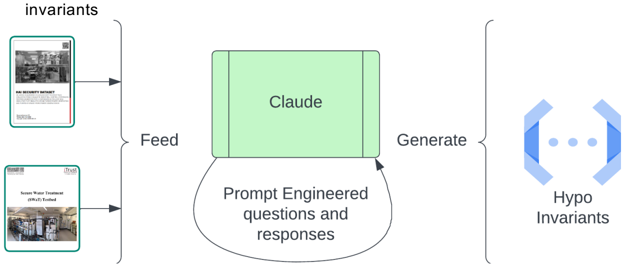
- We are among the first to investigate the use of LLMs in CPS anomaly detection.
- We introduce a novel concept of hypothetical physical invariants that maximizes the capabilities of LLMs and minimizes issues like hallucination.
- A proof-of-concept experiment is conducted using a public CPS dataset to demonstrate the feasibility of the proposed method.

LLM Questions & Responses

Questions	Answers
What physical processes are involved in the testbed?	Boiler process (P1), Turbine process (P2), Water treatment process (P3)
In the boiler process, what major components are used?	main water tank (P1a), heat transfer system (P1b), return water tank (P1c), heating system (P1e), cooling system (P1f)
In the return water tank process, which devices are involved?	LCV01, FCV03, FT03, LIT01
What is the physical relationship among the devices in the form of equations?	$d(LIT01)/d(t) = par1 * (LCV01) - par2 * (FCV03)$

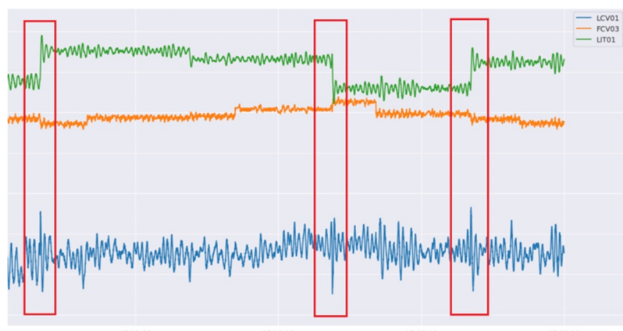
Hypo-invariant Generation

The HAI and SWaT document is input into the pre-trained language model to provide context, followed by the application of prompt engineering techniques, which are used to support the LLM to generate hypothetical invariants



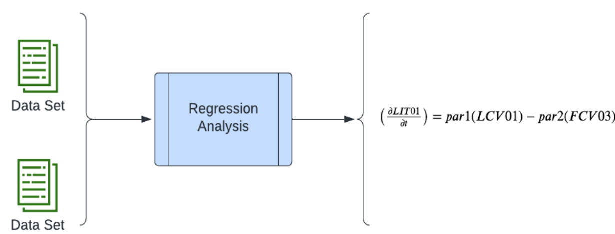
Empirical Study using Line Charts

Data segments from the HAI testbed dataset are identified and collected.



Hypo-invariant Testing

Doing regression analysis with the collected data segments of the sensors to generate some physical formula with relational variables



Data Verification

Values of par1 & par2 on different data segments are gathered from the empirical study of the dataset with the help of regression analysis.

Segment	par1	par2
Seg 1	0.069647	-0.048429
Seg 2	0.092854	-0.182995
Seg 3	0.162843	-0.029161
Seg 4	0.081419	-0.104186
Seg 5	0.080839	-0.037304
Seg 6	0.065190	-0.104206

Conclusion

- **The Potential of LLMs**: Our work states that large language models can deduce hypothetical physical invariants from CPS testbed specifications, enhancing anomaly detection.
- **Empirical Support**: Preliminary results from experiments on the HAI dataset support the proposed method's feasibility.

References

1. Anthropic. Introducing claude.ai. https://www.anthropic.com/, 2023.
2. OpenAI. Introducing chatgpt. https://openai.com/blog/chatgpt, 2022.
3. Giraldo, J., Urbina, D., Cardenas, A., Valente, J., Faisal, M., Ruths, J., Tippenhauer, N. O., Sandberg, H., and Candell, R. A survey of physics-based attack detection in cyber-physical systems. ACM Comput. Surv. 51, 4 (jul 2018).
4. Quinonez, R., Giraldo, J., Salazar, L., Bauman, E., Cardenas, A., and Lin, Z. {SAVIOR}: Securing autonomous vehicles with robust physical invariants. In 29th USENIX Security Symposium (USENIX Security 20) (USA, aug 2020), USENIX Association, pp. 895–912
5. Shin, H.-K., Lee, W., Yun, J.-H., and Min, B.-G. Two ICS Security datasets and anomaly detection contest on the hil-based augmented ICS testbed. In Cyber Security Experimentation and Test Workshop (New York, NY, USA, 2021), CSET '21, Association for Computing Machinery, p. 36–40.