A Preliminary Study on an Intrusion Detection Method using Large Language Models in Industrial Control System

[1]SeokHyun Ann, [1]Seong-je Cho, [2]Hongeun Kim

[1]DanKook University, [2]Dongguk University

1. Introduction

❖ Motivation

- As Operational Technology(OT) and Industrial Control Systems (ICS) evolve, their devices are moving beyond closed networks and connecting to external networks.
- Internet-accessible OT/ICS devices are increasingly exposed to cybersecurity threats.

❖ Objectives

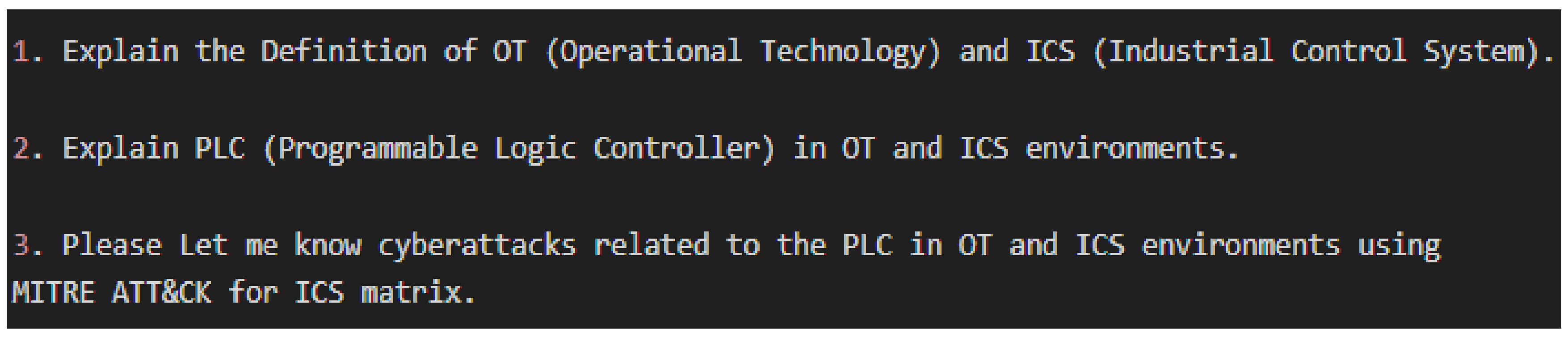
- We propose an LLM-based method to detect intrusion and cyberattacks in OT/ICS environments.
- Instead of detection method based on signatures such as IP, port, and protocol, we use adversary tactics and techniques gathered from MITRE ATT&CK for ICS matrix and open-source data.

2. Issue

❖ Problem

- LLMs sometimes generate incorrect answers because they do not learn based on all the latest facts or evidence.

	GPT - 3.5	GPT - 4	GPT - 4o
Definition of OT/ICS	O	O	O
Definition of PLC	O	O	O
Cyberattacks related to PLC	X	X	X



```
1. Explain the Definition of OT (Operational Technology) and ICS (Industrial Control System).

2. Explain PLC (Programmable Logic Controller) in OT and ICS environments.

3. Please Let me know cyberattacks related to the PLC in OT and ICS environments using
MITRE ATT&CK for ICS matrix.
```

Fig 1. Examples of questions given to an LLM

3. LLM-based Intrusion Detection Model for ICS

Figure 1 shows a method to a LLM for intrusion detection in OT/ICS environments based on MITRE ATT&CK ICS framework.

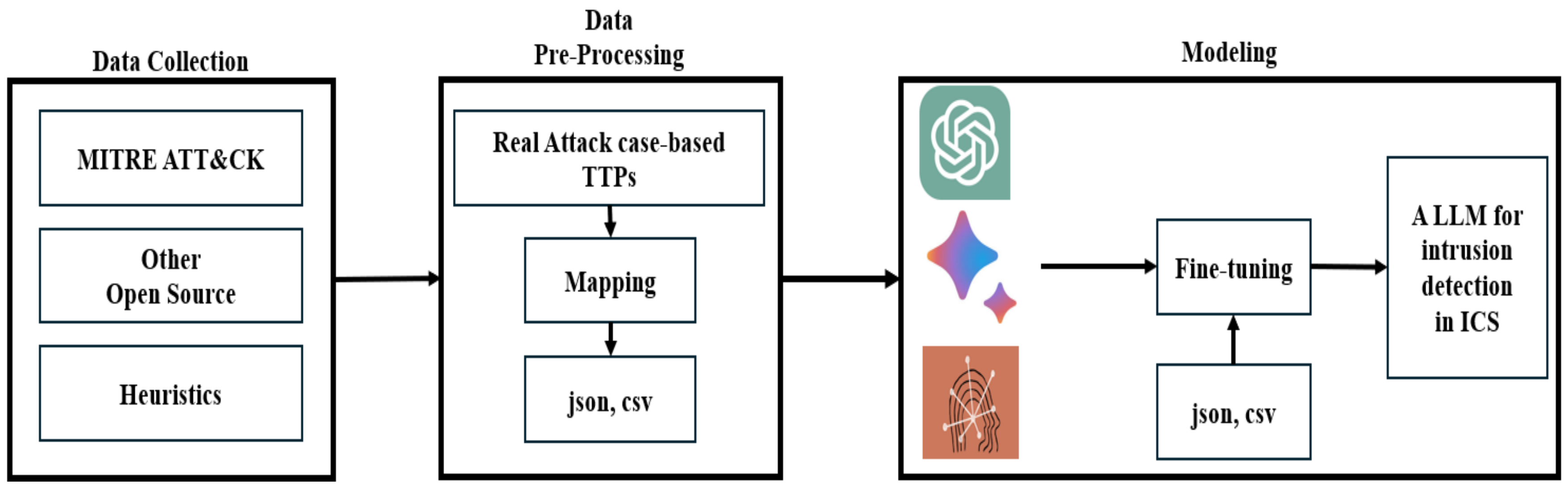
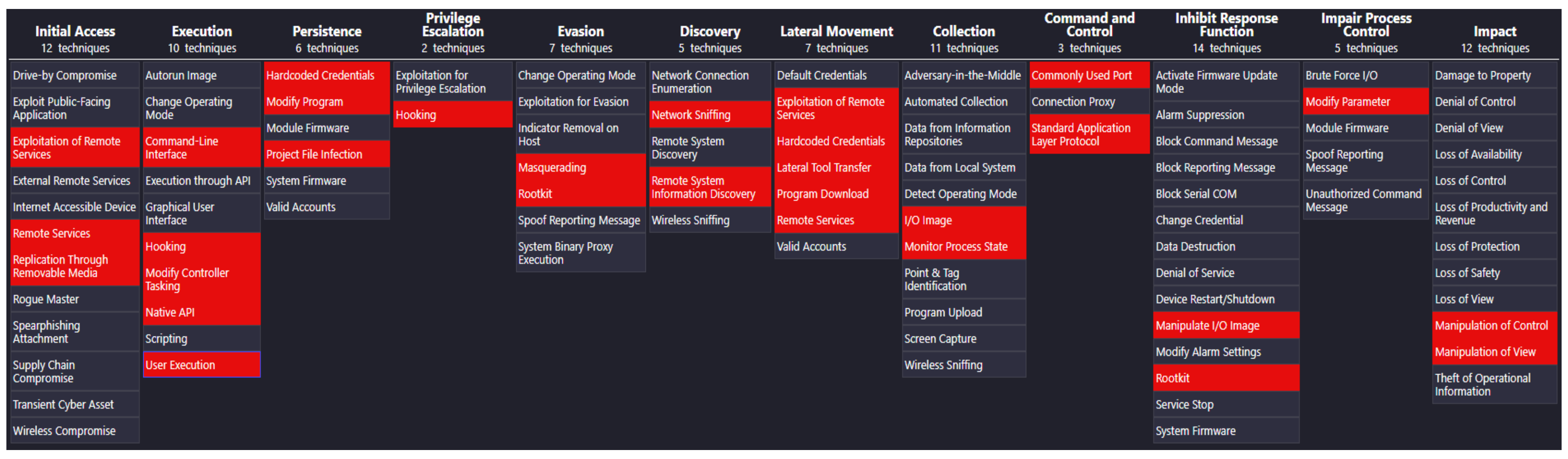


Fig 2. A Structure to Build a LLM for Intrusion Detection in OT/ICS

❖ Data Collections

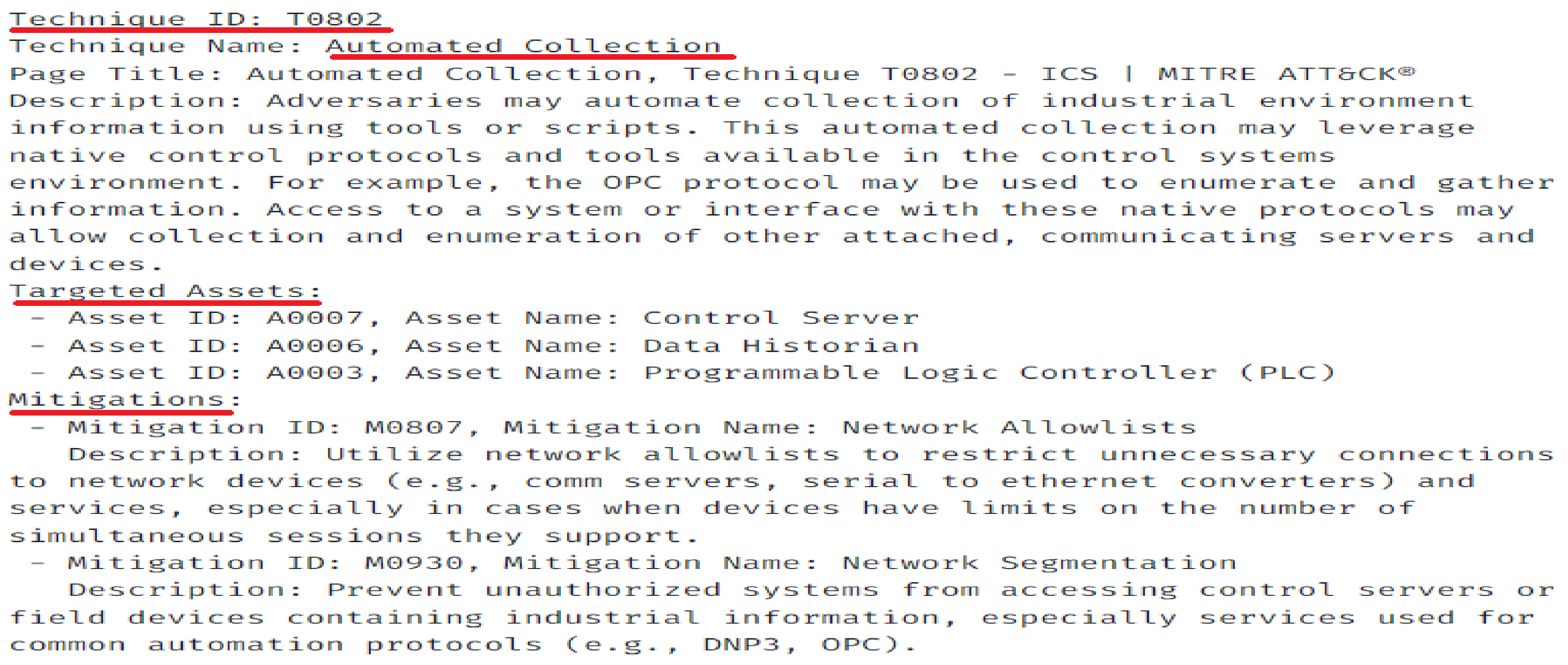
- Collecting information on MITRE ATT&CK for ICS matrix reflecting real-world attacks using Python scripts.
- Collecting other information related to adversary tactics and techniques.

Real Attack case	Used adversary techniques
Stuxnet	22 techniques including Hooking, Rootkit, etc.
PLC-Blaster	5 techniques including DoS, Native API , etc.



Initial Access 12 techniques	Execution 10 techniques	Persistence 6 techniques	Privilege Escalation 2 techniques	Evasion 7 techniques	Discovery 5 techniques	Lateral Movement 7 techniques	Collection 11 techniques	Command and Control 3 techniques	Inhibit Response Function 14 techniques	Impair Process Control 5 techniques	Impact 12 techniques
Drive-by Compromise	Autorun Image	Hardcoded Credentials	Exploitation for Privilege Escalation	Change Operating Mode	Network Connection Enumeration	Default Credentials	Adversary-in-the-Middle	Commonly Used Port	Activate Firmware Update Mode	Brute Force I/O	Damage to Property
Exploit Public-Facing Application	Change Operating Mode	Modify Program	Hooking	Exploitation for Evasion	Network Sniffing	Exploitation of Remote Services	Automated Collection	Connection Proxy	Alarm Suppression	Modify Parameter	Denial of Control
Exploitation of Remote Services	Command-Line Interface	Module Firmware		Indicator Removal on Host	Remote System Discovery	Hardcoded Credentials	Data from Information Repositories	Standard Application Layer Protocol	Block Command Message	Module Firmware	Denial of View
External Remote Services	Execution through API	Project File Infection		Masquerading	Remote System Information Discovery	Lateral Tool Transfer	Data from Local System		Block Reporting Message	Spoof Reporting Message	Loss of Availability
Internet Accessible Device	Graphical User Interface	System Firmware		Rootkit	Wireless Sniffing	Program Download	Detect Operating Mode		Block Serial COM	Unauthorized Command Message	Loss of Control
Remote Services	Hooking	Valid Accounts		Spoof Reporting Message		Remote Services	I/O Image		Change Credential		Loss of Productivity and Revenue
Replication Through Removable Media	Modify Controller Tasking			System Binary Proxy Execution		Valid Accounts	Monitor Process State		Data Destruction		Loss of Protection
Rogue Master	Native API						Point & Tag Identification		Denial of Service		Loss of Safety
Spearphishing Attachment	Scripting						Program Upload		Device Restart/Shutdown		Loss of View
Supply Chain Compromise	User Execution						Screen Capture		Manipulate I/O Image		Manipulation of Control
Transient Cyber Asset							Wireless Sniffing		Modify Alarm Settings		Manipulation of View
Wireless Compromise									Rootkit		Theft of Operational Information
									Service Stop		
									System Firmware		

Fig 3. Tactic and Techniques related to Stuxnet Malware Attack



```
Technique ID: T0802
Technique Name: Automated Collection
Page Title: Automated Collection, Technique T0802 - ICS | MITRE ATT&CK®
Description: Adversaries may automate collection of industrial environment
information using tools or scripts. This automated collection may leverage
native control protocols and tools available in the control systems
environment. For example, the OPC protocol may be used to enumerate and gather
information. Access to a system or interface with these native protocols may
allow collection and enumeration of other attached, communicating servers and
devices.
Targeted Assets:
 - Asset ID: A0007, Asset Name: Control Server
 - Asset ID: A0006, Asset Name: Data Historian
 - Asset ID: A0003, Asset Name: Programmable Logic Controller (PLC)
Mitigations:
 - Mitigation ID: M0807, Mitigation Name: Network Allowlists
   Description: Utilize network allowlists to restrict unnecessary connections
to network devices (e.g., comm servers, serial to ethernet converters) and
services, especially in cases when devices have limits on the number of
simultaneous sessions they support.
 - Mitigation ID: M0930, Mitigation Name: Network Segmentation
   Description: Prevent unauthorized systems from accessing control servers or
field devices containing industrial information, especially services used for
common automation protocols (e.g., DNP3, OPC).
```

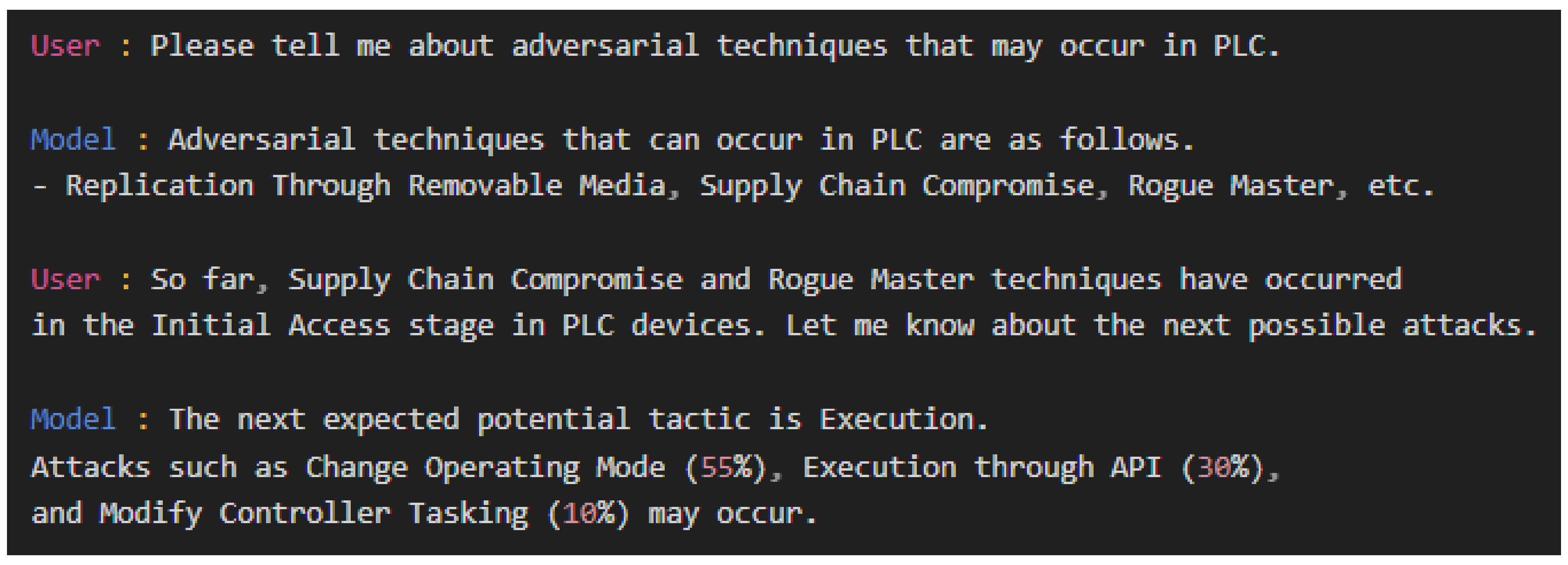
Fig 4. Information of the ‘automated collection’ technique in OT/ICS

❖ Data Pre-processing

- The collected information can be organized based on the MITRE ATT&CK for ICS Matrix.
- It is possible to Identify the tactics and techniques frequently exploited by attackers in OT/ICS environments.
- It processes the information into a format that a LLM can learn from.

❖ Modeling

- By fine-tuning a pre-trained LLM on the dataset, we create a specialized model for detecting and responding to cyber attacks in OT/ICS environments.
- The specialized model can identify attackers’ patterns, or even predict future attacks in OT/ICS.
- As a result, organizations can identify the intent of attackers and minimize the damage from attacks.



```
User : Please tell me about adversarial techniques that may occur in PLC.

Model : Adversarial techniques that can occur in PLC are as follows.
- Replication Through Removable Media, Supply Chain Compromise, Rogue Master, etc.

User : So far, Supply Chain Compromise and Rogue Master techniques have occurred
in the Initial Access stage in PLC devices. Let me know about the next possible attacks.

Model : The next expected potential tactic is Execution.
Attacks such as Change Operating Mode (55%), Execution through API (30%),
and Modify Controller Tasking (10%) may occur.
```

Fig 5. Example of a query and its response between the user and the model

4. Conclusion and Discussion

❖ The differences between OT/ICS environments and traditional IT environments make it difficult to apply existing cyber attack detection solutions.

❖ The signatures used for traditional intrusion detection can be changed, making it difficult to detect and respond to the attacks if attackers evade the detection methods.

❖ We proposed a new LLM-based method to detect cyber attacks in OT/ICS environments using the MITRE ATT&CK for ICS framework.

Dankook University

Computer Security & OS Lab