CSET'22

The picture can't be displayed.

Dataflow-based Control Process Identification for ICS Dataset Development

2022. 8. 8.

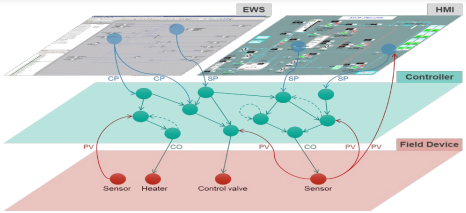
Seungoh Choi, Hyeok-Ki Shin, Woomyo Lee, Jeong-Han Yun, and Byung-Gil Min

CONTENTS

Motivation (1/2)

NSR

Developing a dataset for security research in industrial control systems (ICSs)

Previous work	Problems	Ongoing work
ICS Testbed The picture can't be displayed.	**Biased samples** The representativeness of the control and physical dynamics of the process control system (PCS) is limited.	**Casuality in entire PCS** 
ICS Dataset The picture can't be displayed. The picture can't be displayed. The picture can't be displayed.	**Poor data annotation** Accurately classifying all the affected tags by manually identifying all dependencies on the control logic and physical connectivity is limited.	Generating scenario + Labelling data **ICS Dataset**

Motivation (2/2)

NSR

Three restrictions to identify a casuality of PCS

Limited accessbility

Restrictions on access the control devices during operation

Tangled low-level logic

Practically impossible for a human to utilize a large amount of sequential logic

High vendor dependency

Difference in the control logic expression language from vendors

Available dedicated database + **Data flow graph (DFG)**

Identifying dataflow-based control process for ICS dataset development

General control process

- Digital subsystem

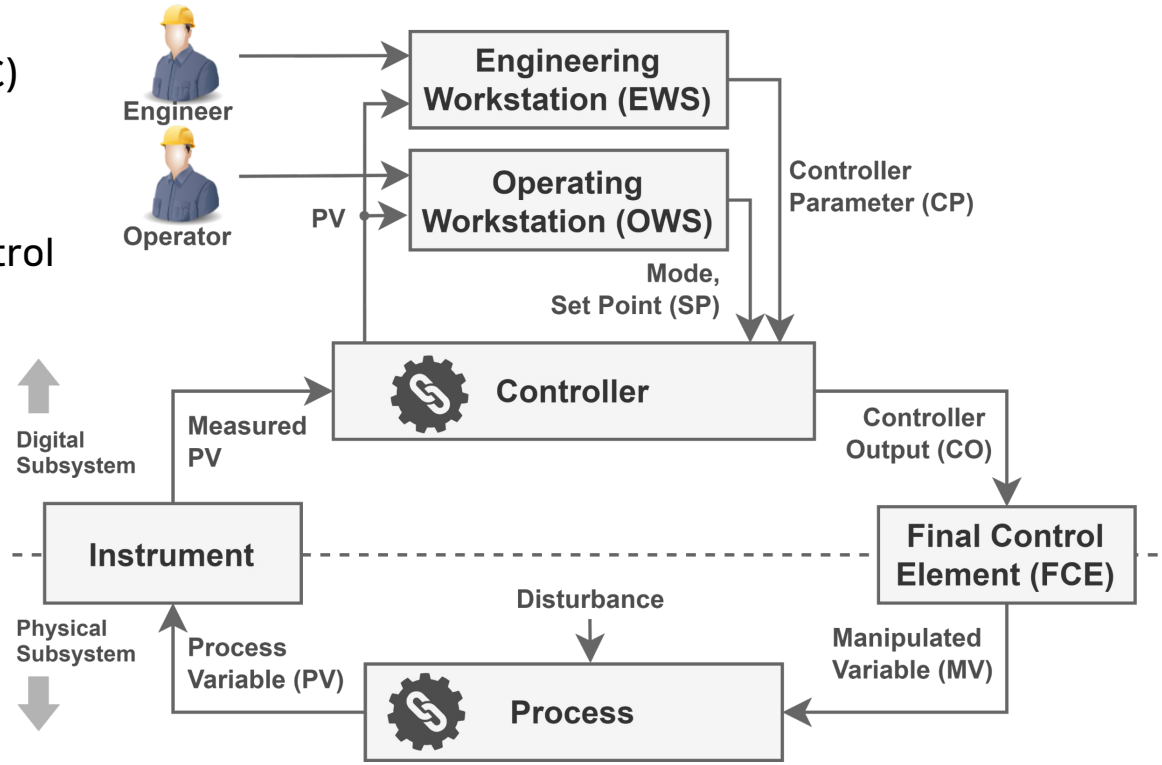
  Management and operation systems: EWS, OWS(HMI), controller (DCS, PLC)

  Variables: SP, Mode, CP, PV, CO

- Physical subsystem

  Field devices: actuators (pumps, control valves) and sensors (transmitters)

  Variables: MV, PV



Gathering control process information

- Tag-related information

  Tags: Names assigned to the control logic variables

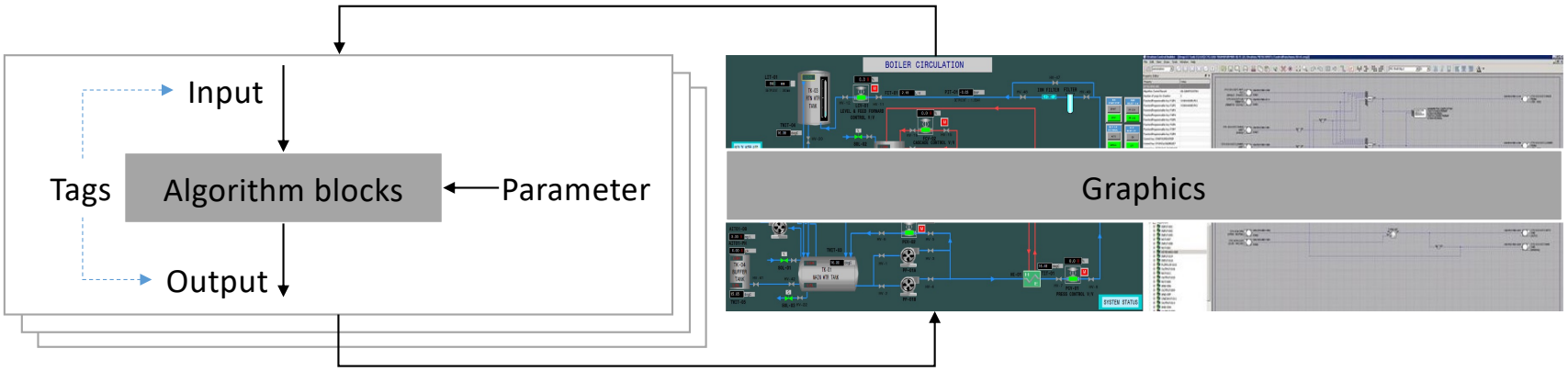
  Algorithm blocks: Constituting I/O tags, and parameters in the control loop

  Graphics: Inculding tags that operators must monitor or that engineers use to tune the system performance

- Information Location

  In general, those information stored in a dedicated database.

  It can be accessed through EWS, historian (HSR), and OPC.



Graph generation

- Identifying nodes

Node type	Digital subsystem I/O		Physical subsystem I/O	
	SP	CP	CO	PV
Target node	○	○	○	
Covert node				○
Initial node	○	○		

- Connectiong edges

Connectivity between digital subsystems

✓ The algorithm block using CP as the input is connected to the EWS node

✓ The algorithm block using SP including operation mode (auto/manual, system on/off) is connected to the HMI node

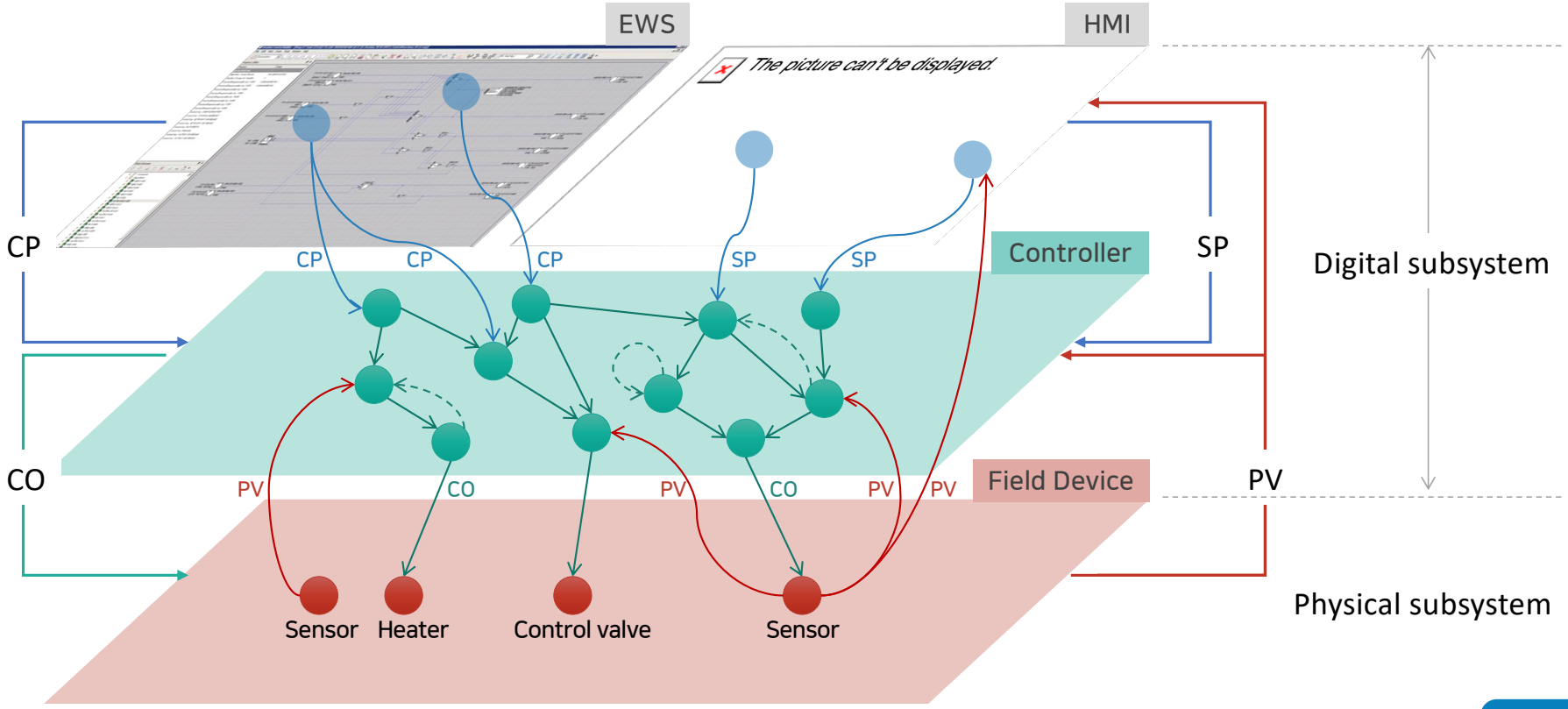
Connectivity between digital and physical subsystems

✓ the physical connection between the controller and the field device corresponds to the algorithim block using PV and CO

✓ PV also corresponds to data flow through the logical connection between the HMI and field devices

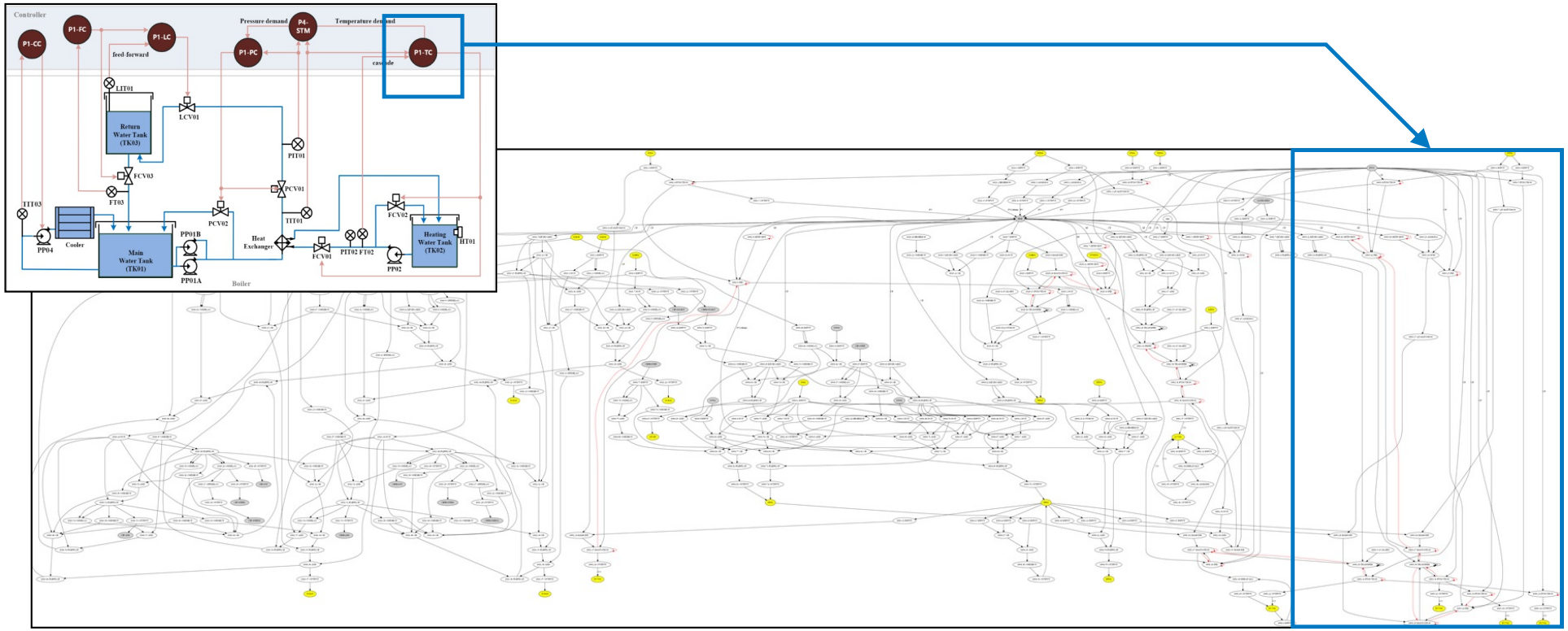
Graph generation

- Example of connectivity between subsystems



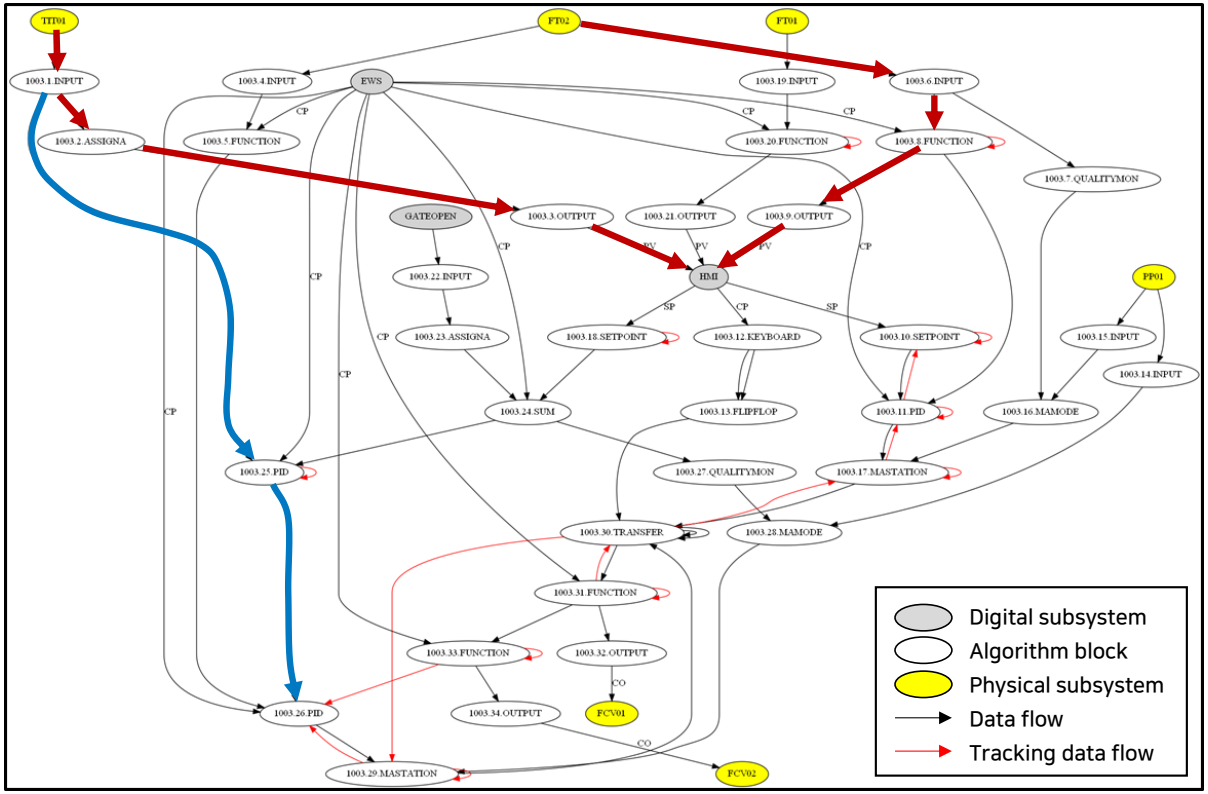
Case I: Graph Generation of Boiler Control System

- Graph for boiler control system and especially represented for one of control loops (P1-TC)



Case I: Graph Generation of Boiler Control System

- Graph for boiler control system and especially represented for one of control loops (P1-TC)



1

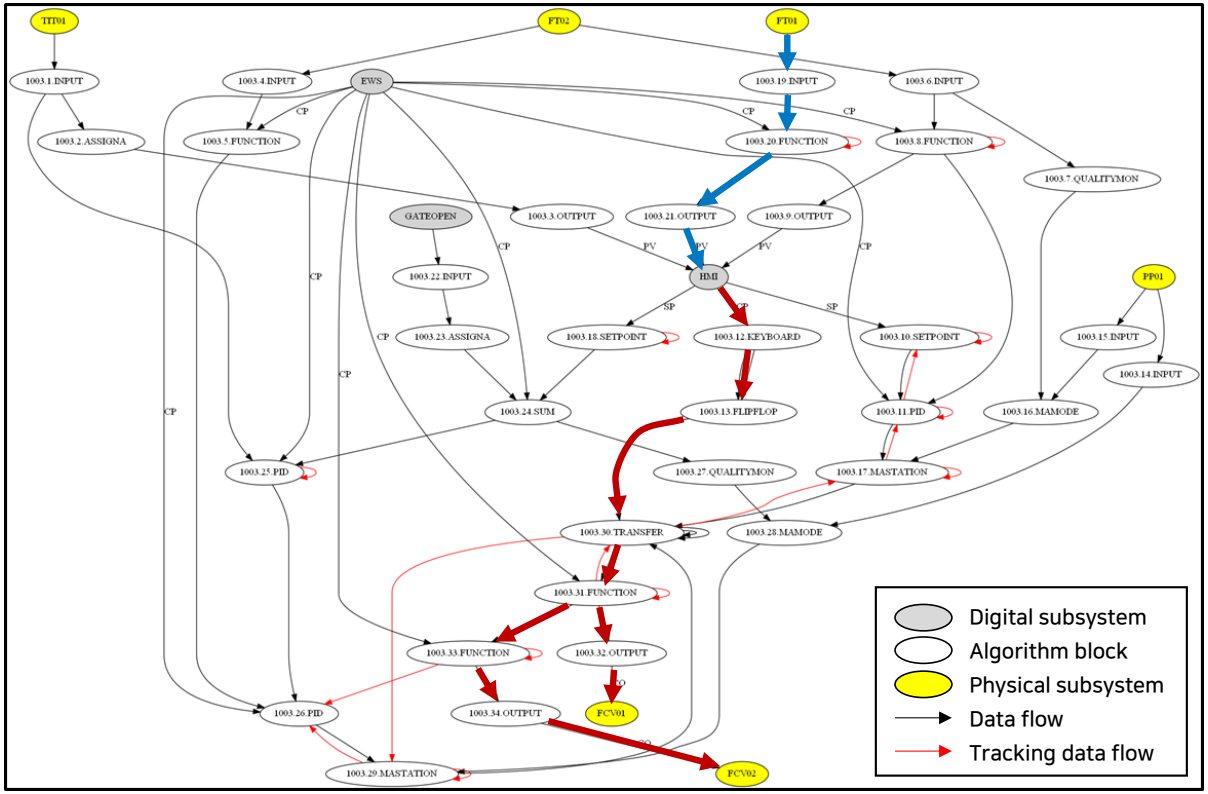
The measured values from the FT02 and TIT01 transmitters on the physical subsystem were obtained through a calibration process (1003.1, 1003.2, 1003.3, 1003.6, 1003.8, and 1003.9) to compensate for the nonlinear distortion of the sensor

2

The heat transfer cascade control was sequentially composed of a temperature control PID node (1003.25) and a flow rate control PID node (1003.26)

Case II: Normal and Attack Scenario Generation

- All paths based on reachability can be identified and configured to generate attack scenario



1

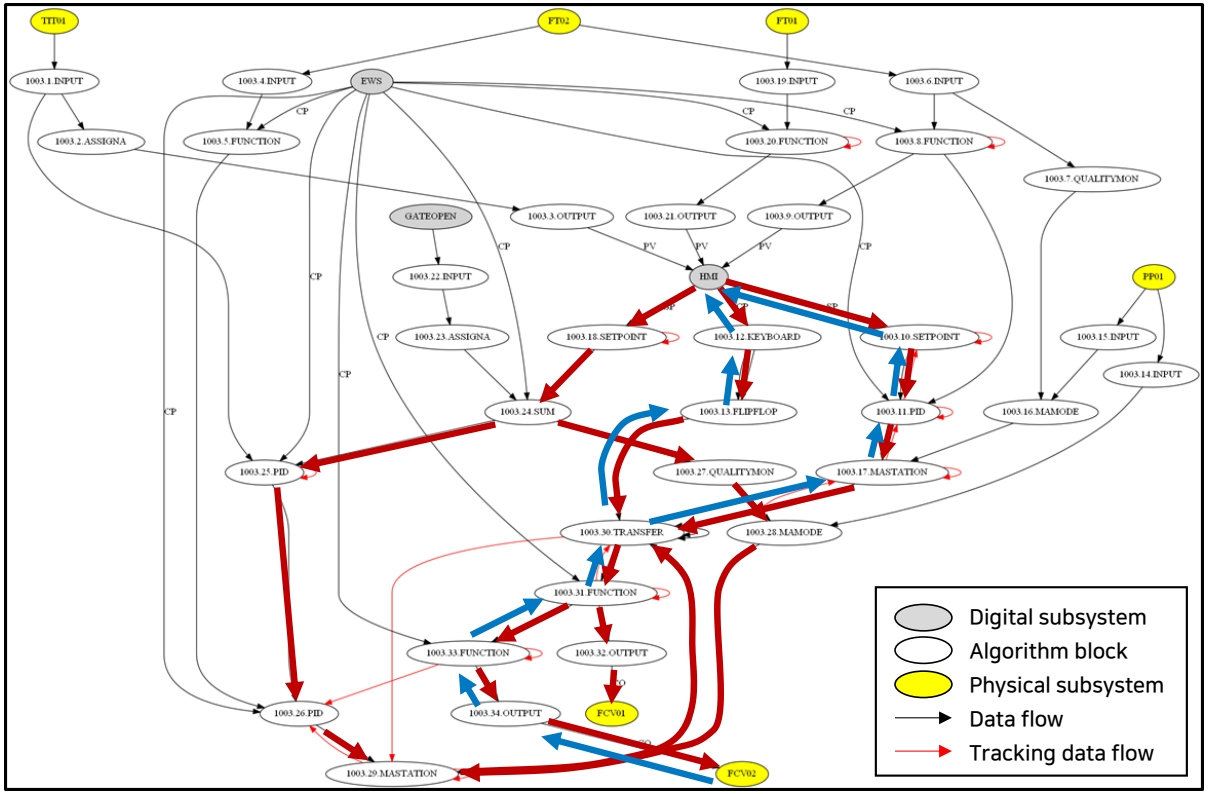
For normal and attack scenario, we assumed operator or attacker manipulate SP for FCV01 or FCV02. By reachability analysis, this scenario is acceptable because there is the path from initial node(HMI) to target node(FCV01 or FCV02).

2

For attack scenario, we assumed attacker manipulate PV for FT01 not to recognized attack. By reachability analysis, this scenario is acceptable because there is the path from initial node(FT01) to target node(HMI).

Case III: Attack Impact and Root Cause Analysis

- All paths based on reachability can be identified and configured to generate attack scenario



1

When HMI is compromised by attacker, attack impact can be identified with attack propargation chain.

2

When backward analysis in the graph, we can provide root-cause information. (e.g. FCV02 Root-cause)

Conclusion

NSR

Available dedicated database + Data flow graph (DFG)

Graph generation to Identify dataflow-based control process for ICS dataset development

Future directions

Dataset deveplotmnet

- **Dataset with informative label**
  - A reasonable dataset can be developed through a precise model based on control process.
  - Informative label including attack impact and root cause information can be provided.

Application coverage

- **All contol processes in HAI testbed**
  - We plan to extend the proposed method horizontally to all control processes of the HAI testbed.
  - And vertically extend it to the physical subsystem by considering the dependency of the system variables.

Development of a dataset of which all control process of the HAI testbed are covered with:

1) Using the proposed method to generate DFG for identify control process
2) then generate data at the HAI testbed as well as informative label for a dataset

The picture can't be displayed.

Q & A

sochoi@nsr.re.kr