1. Description of the experimental environment

The objective of a gas pipeline system is to transmit natural gas from a gate station to a residential dwelling. Its physical environment consists of a gas pump that mimics a gate station, a section of pressure pipeline, three pressure converters, three pressure shut-off valves, and multiple sensors for monitoring pressure, flow rate, and temperature. In particular, the gas pump provides the physical environment with high-pressure gas; the gas pipeline is divided into three sections, namely, the high-pressure line, the medium-pressure line, and the low-pressure line, each of which is controlled by a pair of pressure converters and switches to control the decrease of the gas pressure; multiple sensors are responsible for collecting observations such as the temperature, flow rate, and pressure of the pipeline; and a safety instrumented security system to protect the system's functional safety. In addition, the scenario utilizes two PLCs and an operator control station to control the entire legitimate business process. The business process is as follows:

Turn on the gas pump to generate high pressure; open all valves in turn to construct a gas pipeline delivery path from the gate station to the residences, and turn on all pressure converters to reduce the gas pressure to a usable range for the resident.

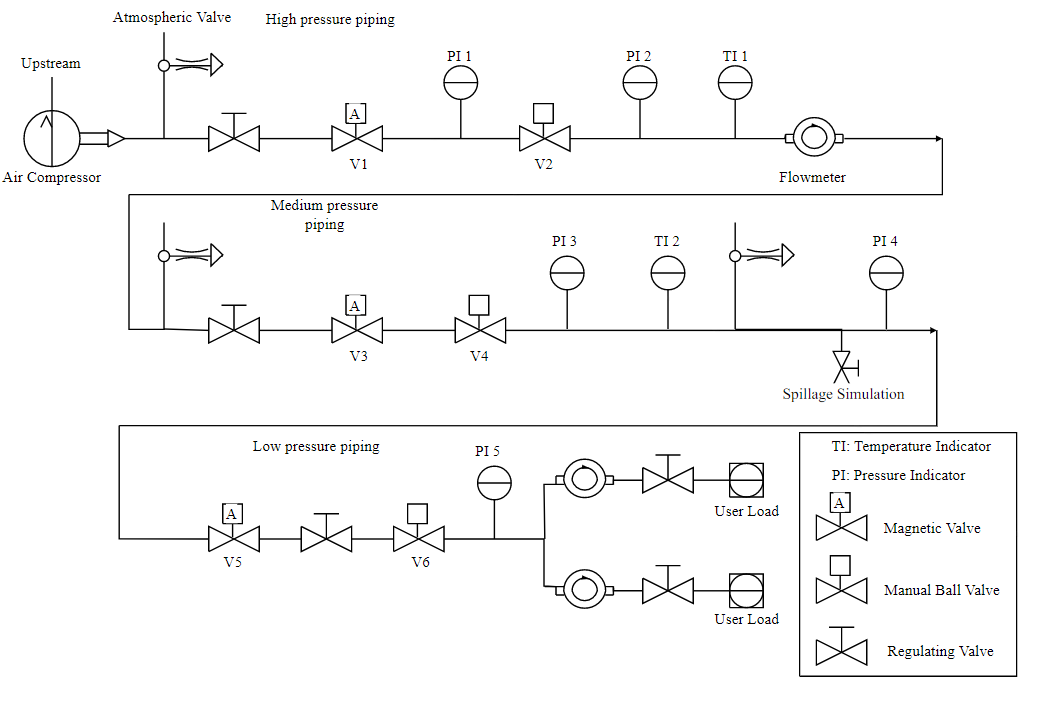


Figure GPN structure

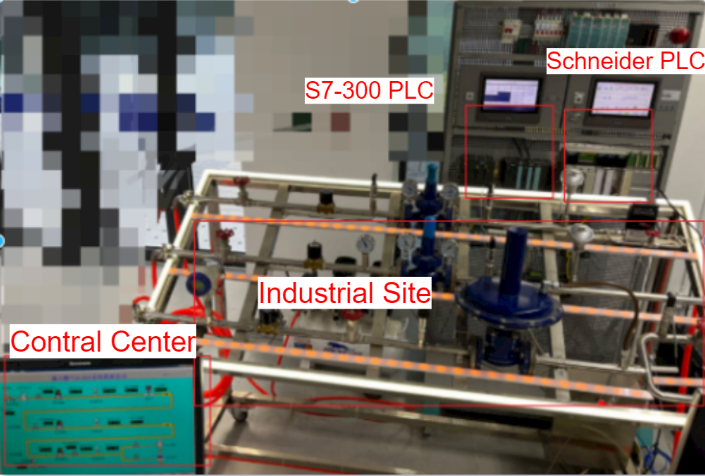


Figure GPN system

2. Introduction to the data set

(1) Data set composition

We design the dataset based on the gas pipeline system. The dataset consists of three parts: attribute information, physical domain information and flow information.

1) Attribute Information

Table Subject attribute

|  |  |  |
| --- | --- | --- |
| Feature Name | Description | Reason |
| User | Entity that issued the visit request | User attributes are used to determine the identity of the accessing entity and its role in the system |
| Role | A collection of all roles created in the system | Roles simplify ICS management and enhance security, and fulfill the principle of segregation of duties |
| Location | The user's geographic location. | The location of industrial control equipment plays a crucial role in authorization assessment. For example, specific tasks can only be performed by subjects within certain designated areas. Geographic information used to identify users. |
| Access Time | User access time. | For example, some industrial control equipment may have restricted opening hours, thereby reducing the likelihood of risk, workload and energy consumption. |
| Operating  System  version | Access the operating system version of the device | Keeping abreast of the operating system versions of industrial control devices ensures timely response to known vulnerabilities and enables more accurate authorization assessment。 |
| Software  version | Access the software version of the device. | Ibid. |
| Geographic  Information | Geolocation information. | Whether a geographic location is under threat and when any threat occurs improves the accuracy of authorization assessments. |
| Frequency | Access frequency. | Monitoring access frequency helps to detect anomalous behavior and enhance the security of the model. |
| Delay | The amount of time spent during a user's visit. | Helps optimize production planning, reduce network congestion risk, and optimize resource utilization |

Table Object attribute

|  |  |  |
| --- | --- | --- |
| Feature Name | Description | Reason |
| Address | Memory address of the object | Maps objects to plc memory addresses for easy management and access. |
| Importance | Importance of the object | Classify objects according to their sensitivity and safety. |

2) Physical domain information

The values of the following sensors and actuators monitored in real time are constructed as physical domain information.

* Gas flow rate
* Temperature
* Medium pressure
* High pressure
* Medium pressure shut-off valve
* High pressure shut-off valve

3) Traffic information

|  |  |  |
| --- | --- | --- |
| **Feature name** | **Description** | **Type** |
| ip.src | Source IP address | Discrete |
| tcp.srcport | Source TCP port | Discrete |
| ip.dst | Destination IP address | Discrete |
| tcp.dstport | Destination TCP port | Discrete |
| frame.protocols | Protocol stack information for data packets | Discrete |
| frame.time\_delta | The time difference between the data packet and the previous data packet | Continuous |
| frame.time\_relative | The time difference between the data packet and the start time of crawling | Continuous |
| frame.time\_delta\_displayed | The time difference between the packet and the previous packet (displayed) | Continuous |
| frame.len | packet length | Discrete |
| eth.src.oui | Manufacturer identification code for the source MAC address | Discrete |
| eth.dst.oui | Manufacturer identification code for the destination MAC address | Discrete |
| tcp.len | TCP packet length | Discrete |
| tcp.ack | Acknowledge number of TCP packets | Discrete |
| tcp.analysis.bytes\_in\_flight | The number of unconfirmed bytes on the network | Discrete |
| tcp.analysis.ack\_rtt | Measure the captured TCP packets and their corresponding ACKs | Continuous |

(2) Dataset setting

1) Set up 50 Users with 45 normal users inside: user1 - user45;

2) Set up 5 abnormal Users: abnormal\_user1 - abnormal\_user5;

Abnormal users: stolen normal users (internal attack)

3) Attack scenarios are divided into

dos, eavesdropping, fake data injection

4) Object resources are divided into

gas flow, temperature, high pressure pressure, medium pressure pressure, low pressure pressure, low pressure cutoff valve, medium pressure cutoff valve, high pressure cutoff valve

5) Role is divided into Temp\_role,Role1-Role3

|  |  |  |  |
| --- | --- | --- | --- |
| Temp\_role | flow, temperature | user30-45 | abnormal\_user4-5 |
| Role1 | Flow, Temperature, Low Pressure, Pressure, Shutoff Valve (Low Pressure) | user1-10 | abnormal\_user1 |
| Role2 | Flow, Temperature Medium Pressure, Pressure, Shutoff Valve (Medium Pressure) | user11-20 | abnormal\_user2 |
| Role3 | flow, temperature High Pressure, Pressure, Shutoff Valve (High Pressure) | user21-30 | abnormal\_user3 |

# 定义客体资源  
class Resource(Enum):  
 gas\_flow = 'DB10.52' # Gas flow  
 temperature\_high = 'DB10.22' # High pressure pipe temperature  
 pressure\_high = 'DB10.2' # High pressure pipeline pressure  
 pressure\_mid = 'DB10.42' # Medium pressure pipeline pressure  
 pressure\_low = '110' # Low pressure pipeline pressure  
 valve\_low = '0' # Shut-off valve low pressure  
 valve\_mid = 'Q0.1' # Cut-off valve medium pressure  
 valve\_high = 'Q0.0' # Cut-off valve high pressure

6) Location

WorkPlace, Home, StrangePlace1-10: of which 1, 2 are contaminated

(3) Data set construction

It is divided into three scenarios: normal user access in normal scenario, abnormal user access in normal scenario and normal user access in three attack scenarios. Among them, normal user access in normal scenario and abnormal user access is 60%:40% totaling 20,000 items, and normal user access in three attack scenarios is 3,000 data.

1) Normal Scenario Normal User Access

Random User1-45 Random guest in authority Location(95% WorkPlace, 4.9% Home, 0.1% StrangePlace) Time(95% 8-20h, 4% 7-8,20-22h, 1% else) OsVer(95% new, 5% old) SwVer(95% new, 5% old)

2) Normal Scenario Abnormal User Access

Random abnormal\_user1 - abnormal\_user5 Random guest Location(4% WorkPlace, 1% Home, 95% StrangePlace) Time(1% 8-20h, 4% 7-8,20-22h, 95% else) OsVer(5% new, 95% old) SwVer( 5% new, 95% old)

3) Normal user access under three attack scenarios

a) dos attack with random User1-45 random guest in privilege Location(95% WorkPlace, 4.9% Home, 0.1% StrangePlace) Time(95% 8-20h, 4% 7-8,20-22h, 1% else) OsVer(95% new, 5% old)

b) Under a theft attack, randomize User1-45 Random guest in privilege Location(95% WorkPlace, 4.9% Home, 0.1% StrangePlace) Time(95% 8-20h, 4% 7-8,20-22h, 1% else) OsVer(95% new, 5% old)

c) Under the fake data injection attack, a random physical domain information is selected to be changed. Random User1-45 Random guest in authority Location(95% WorkPlace, 4.9% Home, 0.1% StrangePlace) Time(95% 8-20h, 4% 7-8,20-22h, 1% else) OsVer(95% new, 5% old)

3. Introduction to the data set format

(1) data1.log

Total 10,000 entries.

Normal Scenario Normal User Access 60% Labeled as 1

Normal scenario abnormal user access 40% access frequency labeled 2 Abnormal behavior and normal behavior 20% each, normal behavior labeled 1, abnormal behavior labeled 2

format: Label (label); Activity (properties); Physics (physical domain information); Net (network information)

Example:

Label;1,Activity;abnormal\_user5|Temp\_role|DB10.52|StrangePlace2|00:00:17|old|old|BadPlace| 0:00:36 |25.176048278808594,Physics;{'gas\_flow': 800.0, 'temperature\_high': 24.467592239379883, 'pressure\_high': 497.97454833984375, 'valve\_high': True, 'pressure\_mid': 234.6643524169922, 'valve\_mid': True},Net;{'ip\_src': '192.168.1.10', 'tcp\_srcport': '56251', 'ip\_dst': '192.168.1.3', 'tcp\_dstport': '102', 'frame\_protocols': 'eth:ethertype:ip:tcp:tpkt:cotp:s7comm', 'frame\_time\_delta': '0.000180000', 'frame\_time\_relative': '9.881918000', 'frame\_time\_delta\_displayed': '0.023853000', 'frame\_len': '85', 'eth\_src\_oui': '3782', 'eth\_dst\_oui': '6939', 'tcp\_len': '31', 'tcp\_ack': '534', 'tcp\_analysis\_bytes\_in\_flight': '31', 'tcp\_analysis\_ack\_rtt': '0.000180000'}

(2) data2.log

Same as above, the difference is using a second host to access (client) changed ip port mac

(3) attack scenario normal user access 5% (external attack)

Label: Dos:3, Steal:4, Tamper:5

Scenario same as normal scenario normal user access, just in external environment for attacking environment.

1) data\_dos.log

Under dos, labeled 3, number 1,000.

2) data\_steal.log

Steal Tagged 4 Quantity 1k entries

3) data\_tamper.log

Tampering: Tagged 5 Qty 1k entries

Physical domain information: With probability 1/6, randomly select one to change

Physical domain information:

Gas Flow, Temperature, Medium Pressure, High Pressure Assume 20% increase of original

Medium pressure shut-off valve, high pressure shut-off valve Change from True to False

*Note: Frequency is the time difference between two visits by the same user, if there is no previous visit is set to 1000.*