UNIVERSITÀ DEGLI STUDI DI VERONA
DEPARTMENT OF COMPUTER SCIENCE

MASTER'S DEGREE IN
COMPUTER SCIENCE AND ENGINEERING

Master Thesis

DECONSTRUCTING THE UNKNOWN: A BLACK BOX APPROACH TO REVERSE ENGINEERING ICS USING GRAPHS AND INVARIANTS

Supervisor:
Prof. **Massimo MERRO**

Co-supervisors:
Prof. **Ruggero LANOTTE**
Marco LUCCHESE

Candidate:
Marco OLIANI
Matr. VR457249

Academic Year 2022/2023

"Gallina vecchia fa buon brodo"

(Valentino Rossi)

Abstract

Bla bla bla

Contents

Chapter 1

Introduction

LOREM ipsum dolor bla bla bla. Ma dove metto l'abstract? Prova di interlinea che direi posso anche andare bene, ma bisogna poi vedere il tutto come si incastra alla fine, in modo da ottenere un bel risultato alla vista.

1.1 Contribution

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

1.2 Outline

Why do we use it? It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout.

The point of using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using 'Content here, content here', making it look like readable English. Many desktop publishing packages and web page editors now use Lorem Ipsum as their default model text, and a search for 'lorem ipsum' will uncover many web sites still in their infancy. Various versions have evolved over the years, sometimes by accident, sometimes on purpose (injected humour and the like).

Where does it come from? Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source. Lorem Ipsum comes from sections 1.10.32 and 1.10.33 of "de Finibus Bonorum et Malorum" (The Extremes of Good and Evil) by Cicero, written in 45 BC. This book is a treatise on the theory of ethics, very popular during the Renaissance. The first line of Lorem Ipsum, "Lorem ipsum dolor sit amet..", comes from a line in section 1.10.32.

The standard chunk of Lorem Ipsum used since the 1500s is reproduced below for those interested. Sections 1.10.32 and 1.10.33 from "de Finibus Bonorum et Malorum" by Cicero are also reproduced in their exact original form, accompanied by English versions from the 1914 translation by H. Rackham.

Chapter 2

Background

2.1 What ICSs are

INDUSTRIAL CONTROL SYSTEMS (ICSs) are information systems used to control industrial processes such as manufacturing, product handling, production, and distribution [1].

ICSs are often found in critical infrastructure facilities such as power plants, oil and gas refineries, and chemical plants.

ICSs are different from traditional IT systems in several key ways. Firstly, ICSs are designed to control physical processes, whereas IT systems are designed to process and store data. This means that ICSs have different requirements for availability, reliability, and performance. Secondly, ICSs are typically deployed in environments that are harsh and have limited resources, such as extreme temperatures and limited power. Thirdly, the protocols and hardware used in ICSs are often proprietary and not widely used outside of the industrial sector.

ICSs are becoming increasingly connected to the internet and other networks, which has led to increased concerns about their security. Industrial systems were not originally designed with security in mind, and many of them have known vulnerabilities that could be exploited by attackers. Additionally, the use of legacy systems and equipment can make it difficult to implement security measures. As a result, ICSs are increasingly seen as a potential target for cyber attacks, which could have serious consequences for the safe and reliable operation of critical infrastructure.

The increasing connectivity of ICSs and the associated security risks have led to a growing interest in the field of ICS security. Researchers and practitioners are working to develop new security technologies, standards, and best practices to protect ICSs from cyber attacks. This includes efforts to improve the security of ICS networks and devices, as well as the development of new monitoring and detection techniques to identify and respond to cyber attacks.

2.2 ICS components

Industrial control systems (ICSs) are composed of several different components that work together to monitor and control industrial processes.

2.2.1 SCADA systems

Supervisory control and data acquisition (**SCADA**) is a system of software and hardware elements that allows industrial organizations to [2]:

- Control industrial processes locally or at remote locations
- Monitor, gather, and process real-time data
- Directly interact with devices such as sensors, valves, pumps, motors, and more through human-machine interface (HMI) software
- Record events into a log file

The SCADA software processes, distributes, and displays the data, helping operators and other employees analyze the data and make important decisions.

SCADA systems are known for their ability to monitor and control large-scale industrial processes, and for their ability to operate over long distances. This makes them well-suited for use in remote locations or for controlling processes that are spread out over a wide area. However, the same features that make SCADA systems so useful also make them vulnerable to cyber attacks.

SCADA systems were not originally designed with security in mind, and many of them have known vulnerabilities that could be exploited by attackers. Additionally, the use of legacy systems and equipment can make it difficult to implement security measures. As a result, SCADA systems are increasingly seen as a potential target for cyber attacks, which could have serious consequences for the safe and reliable operation of critical infrastructure.

To secure SCADA systems, it is important to implement security measures such as network segmentation, secure communication protocols, and access control. Additionally, it is important to monitor SCADA systems for unusual activity and to implement incident response procedures to quickly detect and respond to any security breaches.

2.2.1.1 Scada architecture

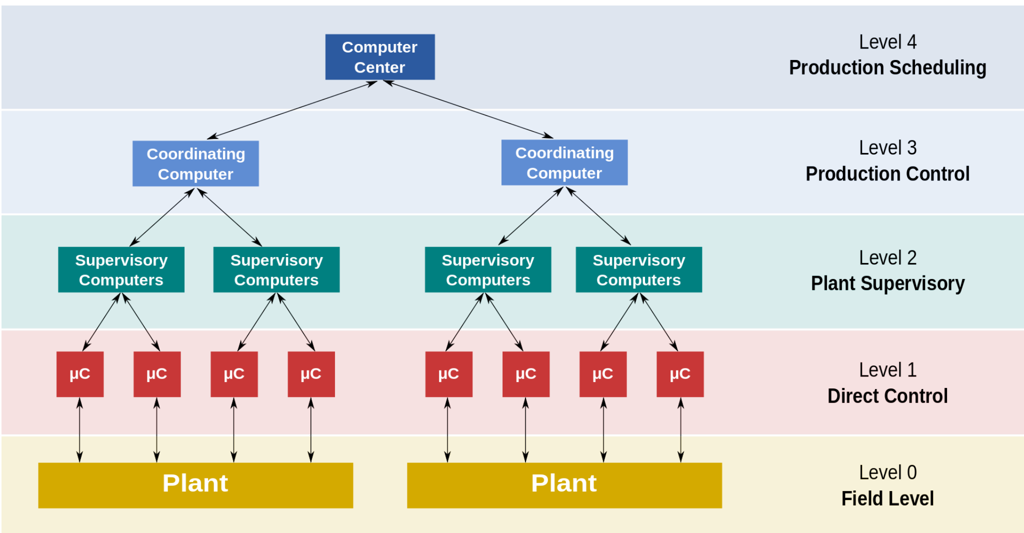


Figure 2.1: SCADA architecture schema

SCADA architecture consists in **five layers** (Figure 2.1):

- Level 0 (**Field Level**): contains **field devices** (2.2.2), or *sensors*.
- Level 1 (**Direct Control**): includes **local or remote controllers** such as **PLCs** (2.2.4) and **RTUs** (2.2.5). Controllers interface directly to the field devices reading data from sensors and sending commands to actuators.
- Level 2 (**Plant Supervisory**): contains computer systems that **collate and store informations** from the previous level and provide a **Human-Machine Interface** (*HMI*, 2.2.6) for operator control.
- Level 3 (**Production Control**): collect and aggregates data from the Plant Supervisory level to generate **reporting** to the Production Scheduling layer.
- Level 4 (**Production Scheduling**): includes business systems (such as ERP systems) used to **manage ongoing processes**.

Production Control level and Production Scheduling level are not directly connected to the process control, but concerned with monitoring production and targets and production scheduling level.

2.2.2 Field devices

Field devices are the **sensors** and **actuators** that are used to collect data from the process and control it. Examples of field devices include temperature sensors, pressure sensors, valves and pumps.

2.2.3 IED

todo

2.2.4 PLC

A *Programmable Logic Controller* (PLC) is a **small and specialized industrial computer** having the capability of controlling complex industrial and manifacturing processes [3].

Compared to relay systems and personal computers, PLCs are optimized for control tasks and industrial environments: they are rugged and designed to withdraw harsh conditions such as dust, vibrations, humidity and temperature: they have more reliability than personal computers, which are more prone to crash, and they are more compact a require less maintenance than a relay system. Furthermore, I/O interfaces are already on the controller, so PLCs are easier to expand with additional I/O modules (if in a rack format) to manage more inputs and ouputs, without reconfiguring hardware as in relay systems when a reconfiguration occours.

PLCs are more *user-friendly*: they are not intended (only) for computer programmers, but designed for engineers with a limited knowledge in programming languages: control program can be entered with a simple and intuitive language based on logic and switching operations instead of a general-purpose programming language (*i.e.* C, C++, ...).

2.2.4.1 PLC Architecture

The basic hardware architecture of a PLC consists of these elements [4]:

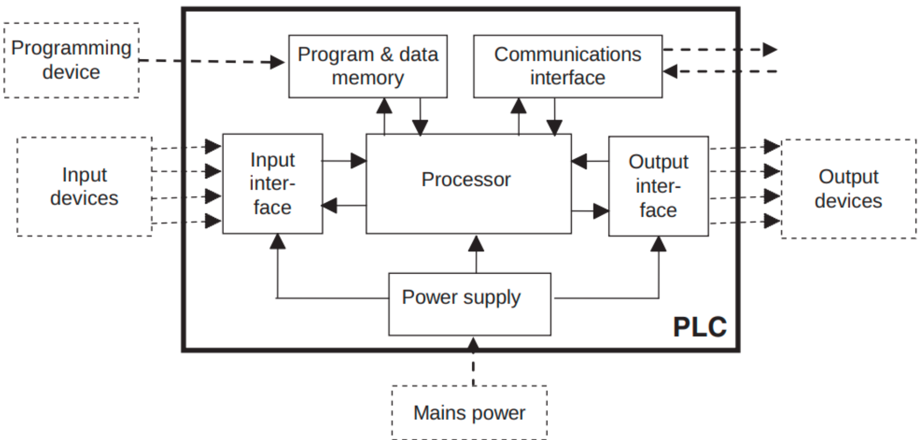


Figure 2.2: PLC architecture

- **Processor unit (CPU)**: contains the microprocessor. This unit interpretes the input signals from I/O modules, executes the control program stored in the Memory Unit and sends the output signals to the I/O Modules. The processor unit also sends data to the Communication interface, for the communication with additional devices.

- **Power supply unit:** converts AC voltage to low DC voltage.
- **Programming device:** is used to store the required program into the memory unit.
- **Memory Unit:** consists in RAM memory and ROM memory. RAM memory is used for storing data from inputs, ROM memory for storing operating system, firmware and user program to be executed by the CPU.
- **I/O modules:** provide interface between sensors and final control elements (actuators).
- **Communications interface:** used to send and receive data on a network from/to other PLCs.

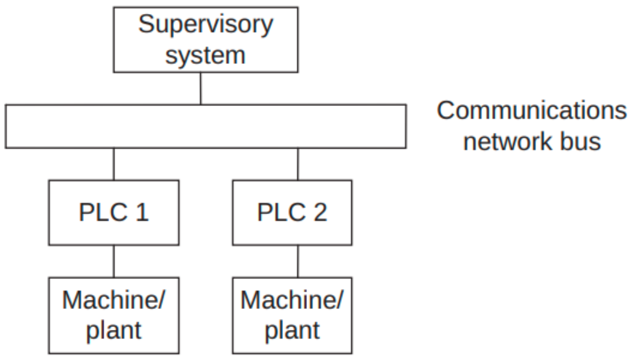


Figure 2.3: PLC communication schema

2.2.4.2 PLC Programming

Two different programs are executed in a PLC: the **operating system** and the **user program**.

The operating system tasks include executing the user program, managing memory areas and the *process image table* (memory registers where inputs from sensors and outputs for actuators are stored).

The user program needs to be uploaded on the PLC via the programming device and runs on the process image table in *scan cycles*: each scan is made up of three phases [5]:

1. reading inputs from the process images table
2. execution of the control code and computing the physical process evolution
3. writing output to the process image table to have an effect on the physical process. At the end of the cycle, the process image table is refreshed by the CPU

Standard PLCs **programming languages** are basically of two types: **textuals** and **graphicals**. Textual languages include languages such as *Instruction List* (IL) and *Structured Text* (ST), while *Ladder Diagrams* (LD), *Function Block Diagram* (FBD) and *Sequential Function Chart* (SFC) belong to the graphical languages.

Graphical languages are more simple and immediate comparing to the textual ones and are preferred by programmers because of their features and simplicity, in particular the **Ladder Logic programming** (see Figure 2.4 for a comparison).

2.2.4.3 PLC Security

PLCs were originally designed to operate as closed systems, not connected and exposed to the outside world via communication networks: the question of the safety of these systems, therefore, was not a primary aspect. The advent of Internet has brought undoubted advantages, but has introduced problems relating to the safety and protection of PLCs from external attacks and vulnerabilities.

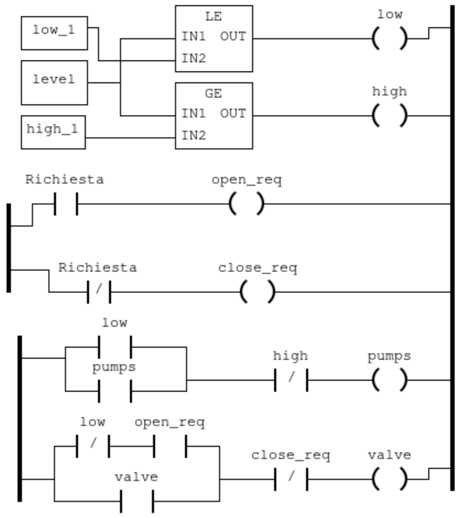
```
PROGRAM PLC1
  VAR
    level AT %IW0 : INT;
    Richiesta AT %QX0.2 : BOOL;
    request AT %IW1 : INT;
    pumps AT %QX0.0 : BOOL;
    valve AT %QX0.1 : BOOL;
    low AT %MX0.0 : BOOL;
    high AT %MX0.1 : BOOL;
    open_req AT %MX0.3 : BOOL;
    close_req AT %MX0.4 : BOOL;
    low_1 AT %MW0 : INT := 40;
    high_1 AT %MW1 : INT := 80;
  END_VAR
  VAR
    LE3_OUT : BOOL;
    GE7_OUT : BOOL;
  END_VAR

  LE3_OUT := LE(level, low_1);
  low := LE3_OUT;
  GE7_OUT := GE(level, high_1);
  high := GE7_OUT;
  open_req := Richiesta;
  close_req := NOT(Richiesta);
  pumps := NOT(high) AND (low OR pumps);
  valve := NOT(close_req) AND (open_req AND NOT(low) OR valve);
END_PROGRAM


CONFIGURATION Config0

  RESOURCE Res0 ON PLC
    TASK task0(INTERVAL := T#20ms,PRIORITY := 0);
    PROGRAM instance0 WITH task0 : PLC1;
  END_RESOURCE
END_CONFIGURATION
```

(a) Example of ST programming



(b) Example of Ladder Logic

Figure 2.4: Comparison between ST language and Ladder Logic

Indeed, a variety of different communication protocols used in ICSs are designed to be efficient in communications, but do not provide any security measure i.e. confidentiality, authentication and data integrity, which makes these protocols vulnerable against many of the IT classic attacks such as *Replay Attack* or *Man in the Middle Attack*.

Countermeasures to enhance security in PLC systems may include [6]:

- protocol modifications implementing **data integrity**, **authentication** and **protection** against *Replay Attacks*
- use of *Intrusion Detection and Prevention Systems* (IDP)
- creation of *Demilitarized Zones* (DMZ) on the network

In addition to this, keeping the process network and Internet separated, limiting the use of USB devices among users to reduce the risks of infections, and using strong account management and maintenance policies are best practices to prevent attacks and threats and to avoid potential damages.

2.2.5 RTU

Remote Terminal Units (RTUs) are computers with radio interfacing similar to PLCs: they transmit telemetry data to the control center or to the PLCs and use messages from the master supervisory system to control connected objects [7].

The purpose of RTUs is to operate efficiently in remote and isolated locations by utilizing wireless connections. In contrast, PLCs are designed for local use and rely on high-speed wired connections. This key difference allows RTUs to conserve energy by operating in low-power mode for extended periods using batteries or solar panels. As a result, RTUs consume less energy than PLCs, making them a more sustainable and cost-effective option for remote operations.

Industries that require RTUs often operate in areas without reliable access to the power grid or require monitoring and control substations in remote locations. These include telecommunications, railways, and utilities that manage critical infrastructure such as power grids, pipelines, and water treatment facilities. The advanced technology of RTUs allows these industries to maintain essential services, even in challenging environments or under adverse weather conditions.

2.2.6 HMI

The *Human-Machine Interface* (HMI) is the hardware and software interface that operators use to monitor the processes and interact with the ICS.

An HMI shows the operator and authorized users information about system status and history; it also allows them to configure parameters on the ICS such as set points and, send commands and make control decisions [8].

The HMI can be in the form of a physical panel, with buttons and indicator lights, or PC software.

2.2.7 Cybersecurity components

Cybersecurity components, as seen in section 2.2.4.3 about PLCs security, are used to protect ICSs from cyber threats and vulnerabilities. They can include firewalls, *Intrusion Detection and Prevention systems* (IDP), and *Security Information and Event Management* (SIEM) systems.

2.2.8 Communication Networks

Communication Networks are the networks that are used to connect the different components of the ICS and allow them to communicate with each other. Communication networks can include wired and wireless networks, such as Ethernet/IP, Modbus, and DNP3 (see Section 2.3).

2.3 ICS Communication Protocols

2.3.1 Modbus

2.3.2 Ethernet/IP

2.3.3 Other Protocols

Chapter 3

Related work

Chapter 4

State of the Art: Presentation of Ceccato et al. Work

4.1 Paper's Goal

4.2 Process Analysis Steps

4.2.1 Scanning of the System and Graph Analysis

4.2.2 Businness Process Analysis

4.2.3 Invariants Analysis

4.3 Application to a Simulated Testbed

4.4 Limitations

Chapter 5

Proposal to Improve Ceccato et al. Work

5.1 Improving Pre-processing

5.2 Improving Graph Analysis

5.3 Improving Invariants Analysis

5.3.1 Automatic Detection of Actuators and Sensors

5.3.2 Invariants Generation

5.4 Obtaining Extra Information on the Runtime Evolution of the Physical Process

5.5 Improving Businness Process Analysis

Chapter 6

Case study: the iTrust SWaT System

Chapter 7

Application to the iTrust SWaT System

7.1 Pre-processing

7.2 Graph Analysis

7.2.1 Conjectures About the System

7.3 Invariants Analysis

7.3.1 Actuators Detection

7.3.2 Daikon Analysis and Results Comparing

7.4 Obtaining extra information on the runtime evolution of the physical system

7.5 Businness Process Analysis

Chapter 8

Conclusions

8.1 Discussions

8.2 Guidelines

8.3 Future work

List of Figures

List of Tables

References

[1] *ICS defintion*. URL: https://csrc.nist.gov/glossary/term/industrial_control_system (visited on 04/06/2023).

[2] *What is SCADA?* URL: https://www.inductiveautomation.com/resources/article/what-is-scada (visited on 04/05/2023).

[3] *PLC defintion*. URL: https://csrc.nist.gov/glossary/term/programmable_logic_controller (visited on 04/06/2023).

[4] William Bolton. *Programmable Logic Controllers, 6th edition*. Newnes, 2015, pp. 7–9.

[5] M. Ceccato, Y. Driouich, R. Lanotte, M. Lucchese, and M. Merro. "Towards Reverse Engineering of Industrial Physical Processes". In: CPS4CIP@ESORICSAt 2022 (Copenhagen, Denmark, Sept. 30, 2022). 2022.

[6] H. S. G. Pussewalage, P. S. Ranaweera, and V. Oleshchuk. "PLC security and critical infrastructure protection". In: 2013 IEEE 8th International Conference on Industrial and Information Systems (Dec. 2013). 2013. DOI: https://dx.doi.org/10.1109/ICIInfS.2013.6731959.

[7] *Remote Terminal Unit*. URL: https://en.wikipedia.org/wiki/Remote_terminal_unit (visited on 04/08/2023).

[8] *HMI defintion*. URL: https://csrc.nist.gov/glossary/term/human_machine_interface (visited on 04/11/2023).