Dissertation Type: Applied

DEPARTMENT OF ENGINEERING MATHEMATICS

Control Logic Based Cyber Attacks in Industrial Control Systems

Laurence Browne

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in the Faculty of Engineering.

Tuesday 26th September, 2023

Supervisor: Dr. Sridhar Adepu

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of MSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Laurence Browne, Tuesday 26th September, 2023

Abstract

This paper investigates the use of machine learning techniques to model the operation of, then create attacks against Cyber Physical Systems (CPS). CPS have a physical aspect, which measure and influence the real world through sensors and actuators, and a cyber aspect, which provide control and monitoring functions.

CPS are traditionally used in industrial process but with the advent of the Internet of Things (IoT) they now proliferate modern life- from home automation systems and autonomous vehicles through to the Critical National Infrastructure such as power grids and public transport networks. As such, understanding the threats posed to these system by emerging technologies is a key area of research for both the public and private sector.

This paper focuses on Secure Water Treatment testbeds (SWaT) as variants of these are present in the majority of industrial process and include a range of components which are common to other processes. Initially a cyber attack methodology was applied to identify aspects of the SWaT which are vulnerable when an 'insider' access level is assumed- that is, full knowledge of, and access to all aspects of the system (including any anomaly detection systems).

A cyber-attack can have multiple facets, investigated here are:

- False anomalies to provoke unnecessary maintenance
- Spoofing control signals to increase wear/ impact system operation
- Hiding/ spoofing error messages to allow components to be damaged

Supporting Technologies

Third-party resources used during the project:

- Development of the SVM and GAN models was performed on a AMD Ryzen CPU and NVIDIA A4000 GPU running Ubuntu V22.04.03. Pycharm Professiononal 2023.1.3 IDE was used for the code development
- Real-Time Modelling of the SWaT was performed in a VMWare Workstation Player virtual machine running Ubuntu V22.04.03. The Mininet network simulation tool with a custom MiniCPS SWaT model. Wireshark was used to analyse network traffic.
- I used the *Pandas* and *Seaborn* public-domian Python Libraries.
- I used LaTeX to format my thesis, via the online service *Overleaf*.

Notation and Acronyms

The following are terms used within the report:

PLC	:	Programable Logic Controller
SWaT	:	Secure Water Treatment Testbed
CPS	:	Cyber Physical System
HMI	:	Human Machine Interface
SCADA	:	Supervisory Control and Data Acquisition
Ethernet/IP	:	Industrial Ethernet
ADC	:	Analogue to Digital Converter
DAC	:	Digital to Analogue Converter
OSI	:	
Sensor	:	
Actuator	:	

Acknowledgements

An optional section, of at most 1 page

TBC.

TBC.

Chapter 1

Introduction

Overview of Cyber PHysical SYstems and the project objectives

The water treatment Cyber Physical System is a collection of sensors which produce continuous signals which are analogous to physical properties of the system such as water level, fluid flow rates and temperature. A Programmable Logic Controller (PLC) monitors these measurements and applies pre-programmed logic to output signals which operate various switches and actuators. The effect of this is that the PLC is a standalone micro-controller which is able to controller an industrial process autonomously. For example, a PLC receiving a low water level signal as an input will send a control signal to a pump or valve which will operate until the input signal is again within range.

The internal logic of the PLC is designed to account for all possible system states and includes parameters such as maximum temperatures, pressures and flow rates. Should the system move outside of these parameters the PLC will take appropriate action and raise an alert. The signals between the PLC and the input and output components are said to be at Level 0 of the OSI model as they either analogue or digital signals in a range suitable for operating switches etc. These signals are typically continuous voltage signals which are converted into a current range between 4 20 mA to avoid transmission losses. This current signal is processed by an external Analogue to Digital Converter (ADC) or Digital to Analogue Converter (DAC) as appropriate as the PLC uses binary encoded values at Level 1 of the OSI model.

A CPS will consist of multiple stages, each performing a distinct process and controlled by it's own PLC. These PLC's communicate with each other in order to pass or request data pertinent to their own stage of the process e.g. request a faster flow of water or indicate a batch is ready to move to the next stage. There are multiple communications protocol's used for this PLC – PLC communication, these are often proprietary to a particular manufacturer but are usually based on the Ethernet/IP industrial Ethernet standard. This configuration is known as Industrial Internet of Things iIOT or Industry 4.0. The updating of the logic within the PLC's is accomplished via a Human Machine Interface which communicates on the same Ethernet/IP ring network as the PLC's. The monitoring of process parameters is performed by Supervisory Control and Data Acquisition (SCADA) sytem which may also play a Historian function and record sytems parameters.

Machine Learning is already applied to these CPS in the form of anomaly detectors. These are usually Support Vector Machine or Deep Neural Networks which have been trained to classify normal and anomalous behaviours. These anomaly detectors may reside on the SCADA system or possinble on a stand-alone processore at PLC level- an example of edge computing. The anomaly detector is often a Support Vector Machine or Deep Neural Network operating as a classifier having been trained on data from normal operation and simulated attacks.

A cyber vulnerability methodology was used to identify vulnerabilities suitable for machine learning techniques to be applied to. The study assumes an 'insider' level of access where the attack has full access to the system- this allows manipulation of the raw signals at level 0 through to the high-level TCP/IP packets which carry the control and monitoring messages.

A SVM Classifier was trained on the sample data to act as a baseline then a Generative Adversarial Network (GAN) was created to generate data for the attack. A GAN consists of a Classifier and a Generator where the error from the classifier is used to improve the generator until it can successfully produce data which can fool the classifier.

This method was applied to the individual Stages of the SWaT system and to system as a whole.

Chapter 2

Literature Review

Introduction to SWat Testbed, iTrust Centre for Cyber Security

2.1 Introduction to SWaT Testbed, iTrust Centre for Cyber Security

The iTrust Centre for Cyber Security operates a Secure Water Treatment Testbed (SWaT) which is used to "support research into the design of secure, public infrastructure" [1]. SWaT is a Cyber Physical System (CPS) so consists of the physical side which implements a process and a cyber side which performs control, monitoring and security. The testbed produces clean water through by using both Ultra Filtration and Reverse Osmosis which is implemented through a six stage, distributed control system which supports wired and wireless communications.

SWaT Stage 1- Raw Water Processing:

SWaT Stage 1- Raw Water Processing: An Alan Bradley Programmable Logic Controllers (PLC) acts as the primary controller for this stage and a second provides redundancy in case of a failure in the primary. Each stage in SWaT uses this dual controller configuration. The naming convention is PLC followed by the stage number, the backup PLC also has 'b' appended to the name so for Stage 1 the PLC's are 'PLC1' 'PLC1b' The PLC manages the flow of raw water from the inlet into the SWaT Stage 2. A motorised valve controls the flow into a storage tank which has four markings (HH, H, L, LL) which indicate maximum to minimum water levels. These markings correspond to numerical water level readings used by the PLC to control in inlet valve and a constant speed pump which feeds water to Stage 2. The numerical water level values used by the PLC are produced by a ultrasonic, water level sensor which uses current signalling (in the 4-20mA range). This analogue signal is digitised for use by the PLC, all tanks within the SWaT use this type of water level sensing. A pH and a Oxidation Reduction Potential (ORP) sensor are present after the constant speed pump and the measurements they produce are sent to the Stage 2 PLC.

SWaT Stage 2- Chemical Dosing/ Pre-Treatment:

Stage 2 controls the addition of three separate chemicals to the water from Stage 1- Sodium Hypochlorite (NaOCl), Hydrochloric Acid (HCl) and Sodium Chloride (NaCl). These chemical are used to balance the pH and ORP of the water as well as disinfection. Dual dosing pumps are used (presumably for redundancy as with the PLCs) and these add the chemicals into the water feed at a rate determined by the Stage 2 PLC.

SWaT Stage 3- Ultrafiltration:

Stage 3 consists of two water tanks with an ultrafiltration unit in between. Tank T301 holds the water from Stage 2 prior to it being pumped through the filter and tank T401 hod the filtered water prior to it entering Stage 4. The ultrafiltration unit consists of progressively fine, micrometer membranes which

remove particulate matter. Stage 3 consists of several motorised valves and pressure sensors which control the flow of water through the filter. The filters become clogged with use so a differential pressure sensor is used to indicate when the pressure across the unit increases- this signal is used by the PLC to instigate a cleaning cycle (back flush of the system). There are additional flow and pressure sensors which monitor the properties of water entering and exiting the filter.

SWaT Stage 4- De-chlorinisation:

In order to prevent oxidation of the membranes within the Reverse Osmosis (RO) unit, chlorine is removed from the water coming from tank T401 using an Ultra Violet de-chlorination unit (UV 401) and also Sodium Bi-sulphate (NaHSO3) from tank T402 if required. An ORP monitor is used to ensure the Chlorine has been removed. The Stage 3 PLC controls this stage of the process.

SWaT Stage 5- Reverse Osmosis:

The RO stage is the most complex as the nano-filters within the RO units (numbered RO 501, RO 502 RO 503) are sensitive to particulates or chlorine which were missed by the previous stages. Water which successfully permeates the RO filters is sent to tank T601, this is final product of the system (the clean water). It is recycled by being sent back to Stage 1. Water that does not permeate through the RO filters is sent to tank T602, this is the reject water and is used to clean the ultrafiltration unit in Stage 3. Stage 5 has motorised valves, flow metres, pH and ORP sensors in order to protect the RO filters. It also has a cartridge filter.

SWaT Stage 6- Backwash:

The SWaT testbed is programmed to initiate a cleaning cycle every 30 minutes which is controlled by the PLC in Stage 6. An additional cleaning process of back-flushing the ultra-filtration unit (to remove particulates from the filters) is instigated when the signal from the differential pressure sensor in Stage 3 exceeds a pre-defined value. The water from these processes is taken from the Stage 5 reject tank (T602) and is expelled from the system after use.

SWaT Network

The control network (or cyber component) is split into two main parts- Layer 0 and Layer 1.

The Layer 0 network is where the PLC headers interface with the actuators and sensors (peripherals) in order to interact with the physical processes. Layer 0 is taken to be at a photon/ electron level i.e. sub-bit level, the PLC's are unable to interpret or produce analogue signals so continuous signals undergo either Analogue to Digital Conversion (ADC) or Digital to Analogue Conversion (DAC) as needed within the Remote Input Output (RIO) unit. The PLC run software for the control logic of these peripherals (via the RIO) and is connected by a ethernet based circular/ ring network to the RIO and backup PLC.

In the Layer 1 Network the PLCs are connected to each other (using the manufactures proprietary protocol), a SCADA system, a network Historian and a Human Machine Interface. These devices are configured by a star network via a central, 24 port switch. This is Layer 1 in the OSI model as it is the physical layer where packets of binary data are passed, rather than the analogue or binary encoded analogue values passed at Layer 1. The protocol used on this Layer 1 network is not TCP/IP but a specific protocol for control systems which is built on top of TCP and allows data such as that for programming/ updating the logic of the PLCs or firmware updates to be passed. This protocol is also used on the ring part of the Level 0 network where the PLCs and RIO communicate (along with other, undefined components which can communicate via ethernet).

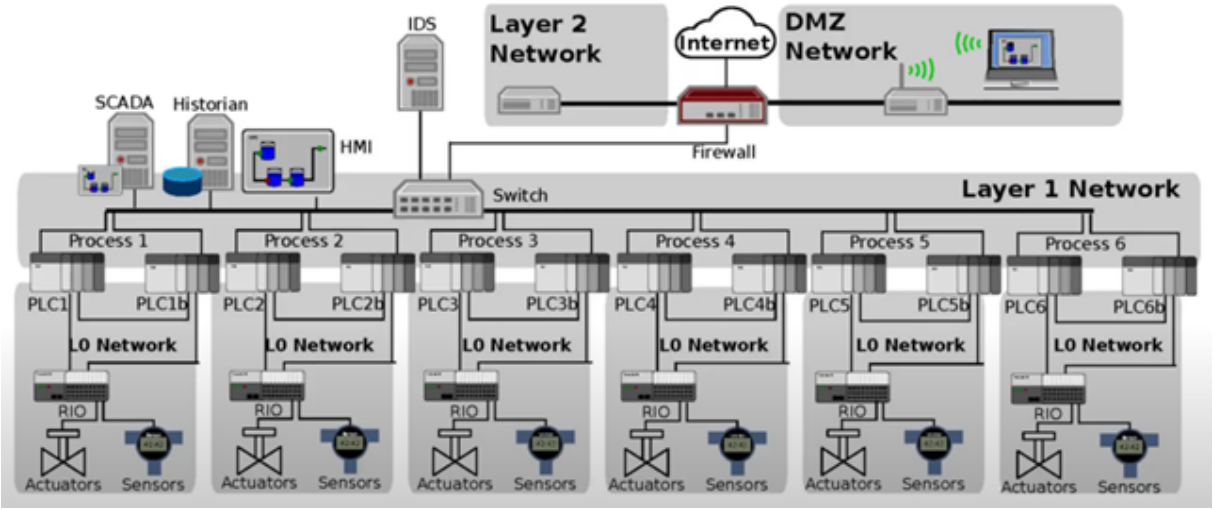
The central switch also used 4 ports to export copies of all the network traffic for use in intrusion detection system development. Both network layers can be changed from wired to wireless communication via a switch on the cabinet for each Stage.

graphicx

2.2 Threat Modeling of Cyber-Physical Systems - A Case Study of a Microgrid System

Shaymaa Mamdouh Khalil a, Hayretdin Bahsi, Henry Ochieng' Dolaa, Tarmo Korõtko, Kieran Laughlinc, Vahur Kotkas

Figure 2.1: SWaT Network Diagram



"Cyber threat modeling is an analytical process that is used for identifying the potential threats against a system" [3]. The article seeks to apply the secure-by-design model from software development to CPS which it states has not been well catered for with current threat models. The paper points out that the CPS are more varied in their makeup than purely software systems so require a input from a wider range of stakeholders such as experts who deal with physical processes. This aligns with the Level 0 in the SWaT testbed whereby the physical aspects of the system are a vector which could be used to attack the system, the paper suggests that software based security models assume the physical aspects of the system (i.e. access to equipment) is secure. The research attempts to develop existing security practices and map these to the IEC 62443 standard which addresses cyber security for automation and control systems. This is implemented in the following 9 stage Threat Modelling Methodology.

Process

Stage 1: Initial Attack Taxonomy Creations Review literature on attacks on similar systems in order to become familiar with the system and it's security issues. Stage 2: Information Systems Assets Identification Identify all system assets whether or not they are within the scope of the threat modelling exercise. In CPS there is likely in two distinct information categories- control and measurement. This could take the form of a system architecture diagram. Stage 3: System Mapping into Data Flow Digagram This stage allows the visualisation of assets which produce/ use data but do not have computing capability so would be outside the scope of conventional cyber threat modelling. Stage 4: Security Context Definition Agrees the physical security assumptions such as who is trusted with admin access and the main threat actors. Stage 5: Trust Boundaries Determination

Stage 6: Threat Elicitation and Attack Taxonomy Update Applies STRIDE to each element in the DFD or to information flows which cross a trust boundary. Stage 7: Threat Consequences Losses Identification Cyber security experts work with system experts to identify real world consequences of each threat. Stage 8: Threat Prioritisation Highest impact threats prioritised. Stage 9: Security Requirements Selection System requirements required in order to counter identified threats.

2.3 Adversarial Attacks and Mitigation for Anomaly Detectors of Cyber-Physical Systems

Chen et al describe the methods used by CPS to identify anomalous behaviour which is indicative of a cyber attack. They describe that typically a CPS has two forms of defence: Firstly, an anomaly detector which is a Machine Learning model (often based on a neural network model) which is trained on the systems physical data. Secondly, rule checkers (or invariant checkers) are used which check values against the acceptable parameters or known relationships between components in the CPS . Chen et al assume a 'white' box level of access to the anomaly detector, that is a full understanding of it's behavious

and accesst to the data it was trained on .It is assumed the rule checker is only a black or grey level of access so it's behaviour must be learnt from the librarian logs etc.

The team 'crafts noise' over the signal between actuators and sensors then use a 'genetic algorithm' to optimise the noise so that both detection systems are deceived to the degree that their classification accuracy is reduced by over 50

The report mentions how attacks on the CPS typically involve spoofing or manipulating the network packets and neural network based detectors are effective at identifying these. This paper seeks to create attack possible when there is 'insider' level access- the attacker knows the anomaly model. The focus of the paper is to create noise which will lead the anomaly detector and rules checkers to misclassify the activity. For example, if the attack can use noise to shrink the difference between the actual value and the predicted value then the anomaly detector will assert more false positives "when a detector misclassifies a real attack as normal behaviour" [2]. Jiaa et al assert that "existing adversarial attacks have limited effectiveness in the presence of rule checkers" but that genetic algorithms based on the white-box gradient based approach can remedy this.

The paper defines a CPS as PLC's which are connected to actuators and sensors which are the interface to the physical world. The PLC run software for the control logic of these peripherals which it is connected to by a circular/ ring network operating at 'Layer 0'. Layer 0 is taken to be at a photon/ electron level- i.e. sub-bit level so continuous or discrete signals. These PLCs are connected to a central SCADA system by a star network operating at layer 1- the physical layer.

It is assumed that rule checkers reside within the PLCs- for example to open a valve using an actuator when a particular sensor value is met. The anomaly detector is assumed to reside on the SCADA system.

The paper describes the two test beds used for the research – the Swat WADI plants that model a water treatment and water distribution plants respectively. The SWAT plant is described as having 68 sensors and actuators in total, a number of these are standby in case of failures and were not cond=sidered in the paper. It is noted that the sensors are typically continuous values and the acutators are discrete. This is understandable as the output of the PLC is likely to a motor controlller or relay which handle things like soft start for motors or gradual closing of valves in order to avoid the water hammer effect (me).

'Our approach is inspired by a white-box gradient-based approach [33],' N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, A. Swami, The limitations of deep learning in adversarial settings, in: 2016 IEEE European Symposium on Security and Privacy (EuroSP), IEEE, 2016, pp. 372–387

graphicx

Figure 2.2: Adversarial attack Diagram

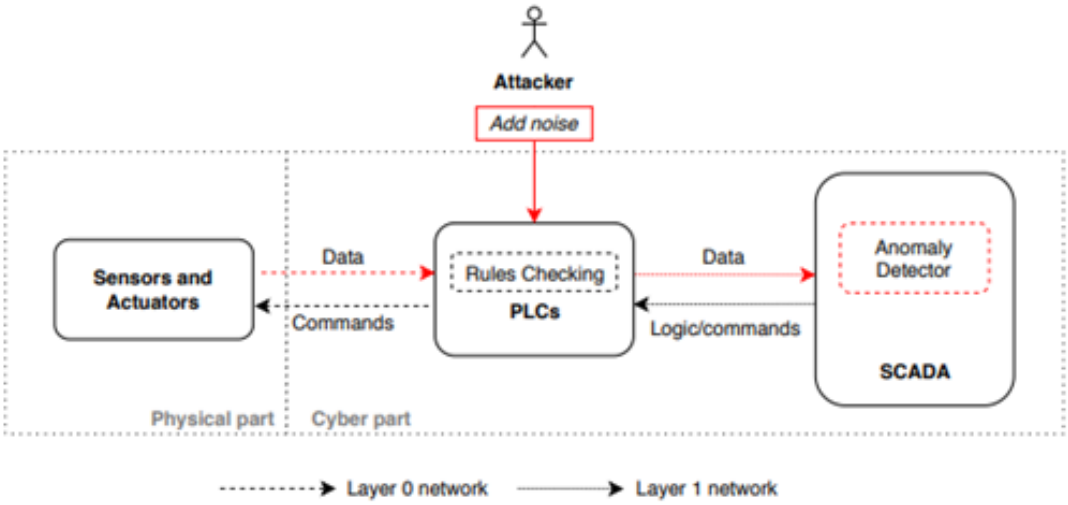


Figure 1: Overview of a cyber-physical system and an adversarial attacker

The SWaT testbed has a historian which records the physical state of the system, this is "a fixed ordering of all the sensor readings and actuator configurations at a particular timepoint" [3]. The report uses the following notation to denote the system state (x) where subscript a and s are used for acutators

and sensors respectively.

graphicx

Figure 2.3: Ststem State Formular

$$x = [x_{a1}, x_{a2}, x_{a3} \ldots x_{s1}, x_{s2}, x_{s3} \ldots]$$

SCADA, Historian and Human Machine Interface workstations sit at higher levels and allow operations such as changing control code/ parameters within the PLCs. The Threat model used is White Box where attacker has access to physical signals at layer 0, full knowledge of the RNN based anomaly detector but can only judge rule checker from Status in the historian. The authors use gradient based methods where by the original attack signals have noise added which is basedon the loss gradient of the RNN. This does not affect the attack but leads to the attack being misclassified.

Bibliography

[1] iTrust Centre in Cyber Security. Introduction to swat testbed, 2016. Accessed: 5 April 2016.

[2] Yifan Jiaa and J. W. C. M. P. S. C. J. S. Y. C. Adversarial attacks and mitigation for anomaly detectors of cyber-physical systems. *International Journal of Critical Infrastructure Protection*, 2021.

[3] Shaymaa Mamdouh Khalil and H. B. H. O. D. T. K. K. L. V. K. Threat modeling of cyber-physical systems - a case study of a microgrid system. *Science Direct*, January 2023. Online; Accessed: [Your Access Date here].

Appendix A

Project Timeline

Proposed breakdown of project elements by week:

Figure A.1: Project Timeline

CPS Project Timeline 24/09/2023

	Percentage Complete	11/09/2023 1	18/09/2023 2	25/09/2023 3	02/10/2023 4	09/10/2023 5	16/10/2023 6	23/10/2023 7	30/10/2023 8	06/11/2023 9	13/11/2023 10	20/11/2023 11	27/11/2023 12	04/12/2023 13	11/12/2023 14	18/12/2023 15
Initial Investigation	100%															
Literature Review	50%			Draft Lit. Review	Lit Review											
Exploratory Data Analysis	30%															
Methods & Tools	30%															
Project Planning	50%			Draft Plan	Plan											
SVM Classifier	0%						Working Classifier									
GAN Classifer	0%								Working Classifier							
GAN Generator	0%										Working Generator					
Attack Evaluation	0%															
Report Writing	0%														Final Report	
Video Presentation	0%															Video Presentation

Milestone Deliverable