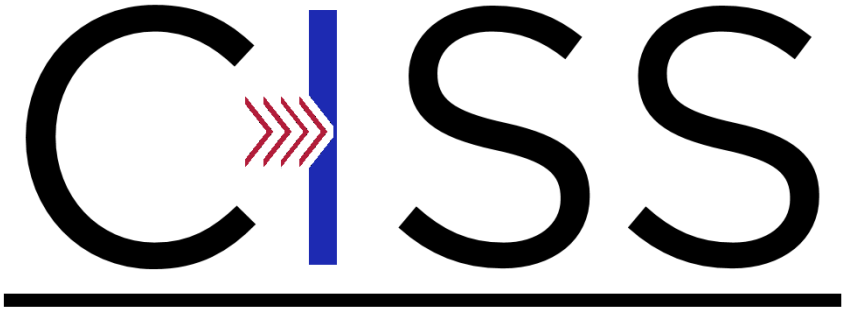
The Third International



Critical Infrastructure Security Showdown 2019

Technical Report

Sponsor	:	National Research Foundation
Date	:	26 – 29 August 2019
Venue	:	SUTD Campus, Building 2, Level 7, 2.705, Secure Water Treatment Testbed
Reported by	:	Beebi Siti Salimah Binte LIYAKKATHALI, Francisco FURTADO, Kandasamy Nandha KUMAR, and Ivan LEE
Organiser	:	iTrust, Centre of Research in Cyber Security

Table of Contents

1 Introduction

The Critical Infrastructure Security Showdown (CISS) 2019 ("Exercise"), conducted in August 2019 at SUTD, was the third run of iTrust's technology assessment exercise. This Exercise was sponsored by the National Research Foundation, Singapore. It consists of three main phases, a setup phase from 5 August to 22 August, a walkthrough phase on 26 August followed by the attack phase from 27 August to 30 August. Red and Blue Teams from academia and industry were invited to participate in this Exercise.

The objectives of the Exercise were to enable:

- Researchers in iTrust to: (1) empirically evaluate defence mechanisms developed in-house against skilled attackers, (2) be exposed to and discover new attack vectors to defend against; and (3) strengthen the existing defence mechanisms;
- red teams to have a unique opportunity to attack the Secure Water Treatment (SWaT), a 5 US gallon/min industrial water treatment testbed. For additional realism, Red Teams were encouraged to enter SWaT's network via the ZyCron Cyber City (ZCC), which simulates a plant operator's enterprise network; and
- blue Teams to showcase their detection capabilities against cyber-attacks.

2 Phases

2.1 Set-Up Phase for Blue Teams

From 5 August to 22 August, each Blue Team was given up to 3-working days to set up their defence mechanism in SWaT. Once all defence mechanisms had been set up, iTrust operated the testbed under "normal operating conditions" for up to 3-days. This gave the defence mechanisms time to learn the plant behaviour.

Four commercial defence systems and six iTrust defence technologies were deployed throughout the entire Exercise. The objective was to detect and raise an alarm when an attack is detected under a realistic setting such as in a water treatment plant. The detectors are designed to detect tampering of plant processes after attackers have successfully obtained network or physical access into the system, beyond traditional network-centric infrastructures such as a firewall, or by breaching physical security. While a traditional intrusion detection system (IDS) may detect intrusions, it is unlikely to be equipped with control strategies to mitigate the impact of cyber-attacks. The blue teams could only install an IDS, without any counter attack or prevention add-ons so as not to disrupt the Red Teams carrying out their attacks. The six defence systems deployed by iTrust were Argus, DAD, GARX, NoisePrint, MLP and HybMonitor.

2.2 Walkthrough Phase

Prior to the Exercise, the seven Red Teams were invited to visit the SWaT testbed to familiarise with SWaT and ZCC setup. No attacks or connections of any sort to SWaT were allowed during this phase. Attendance in this phase was optional and did not affect the team's final score. The Red Teams are: CTF.SG, North Atlantic Treaty Organization Cooperative Cyber Defence Centre of Excellence (NATO CCDCOE), Tower of Hanoi, NARA Institute of Science and Technology (NAIST), Red Alert, ABZB LLC and iTrust Red Team.

2.2.1 ZyCron Cyber City (ZCC)

ZCC is a full-fledge virtual organisation comprising Information Technology (e.g., e-mail server, file server, printer server, CCTV, honeypot and intranet, but no internet access) and Operational Technology (water treatment processes in SWaT), that are meaningfully represented. To make these entities "alive," various types of network traffic were crafted and included in ZCC. As an IT environment, ZCC is not set up with best practices i.e. it is intentionally built with minimum security features and contains vulnerabilities for Red Teams to explore and exploit. A high-level architecture of ZCC is presented in Figure 1.

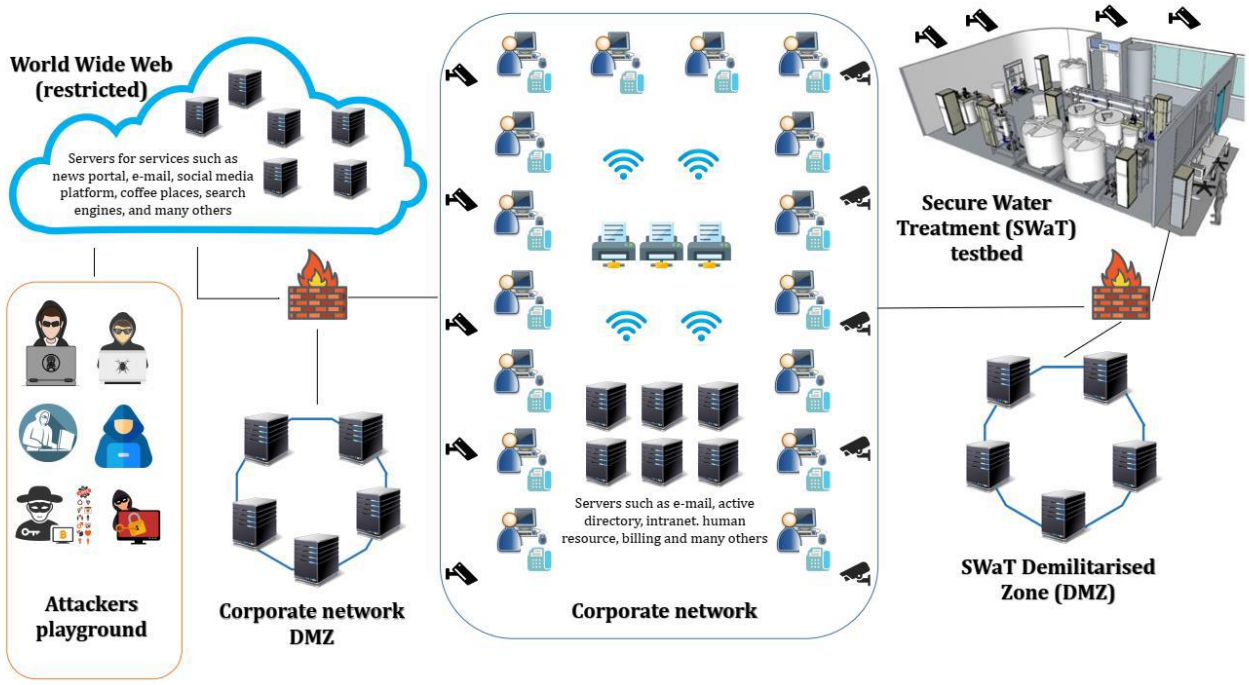


Figure 1: High-level Architecture of ZyCron Cyber City

2.2.2 SWaT Network Architecture

The detailed architectures of SWaT can be found here and here; however, as a means of demonstrating the scale of these testbeds, and framing the discussion of the attacks conducted, SWaT is discussed briefly below.

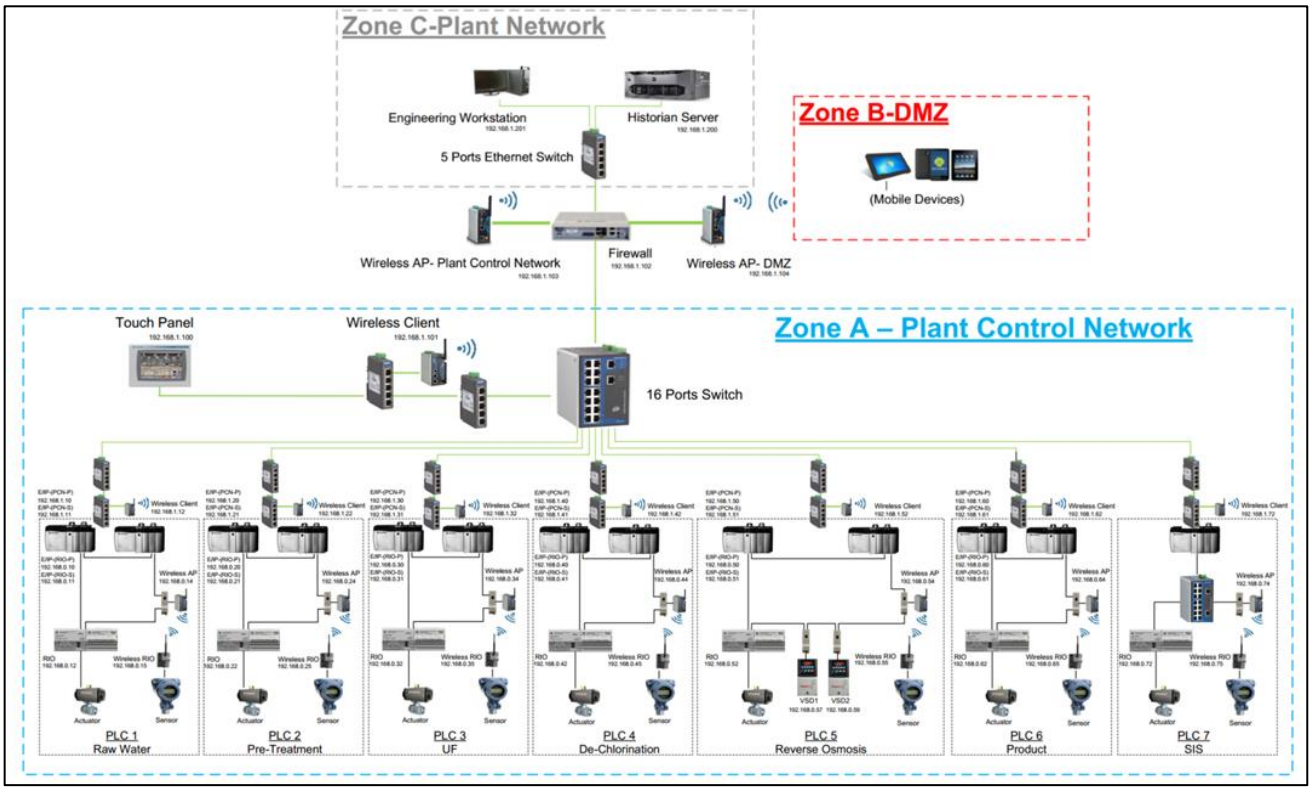


Figure 2: Network Architecture of SWaT

SWaT consists of a modern six-stage process. The process begins by taking in raw water, adding necessary chemicals to it, filtering it via an Ultrafiltration (UF) system, de-chlorinating it by using ultra-violet (UV) lamps, and then feeding it to a Reverse Osmosis (RO) system. A backwash process cleans the membranes in UF using the water produced by RO. The cyber portion of SWaT consists of a layered communications network, Programmable Logic Controllers (PLCs), Human Machine Interfaces (HMIs), Supervisory Control and Data Acquisition (SCADA) workstation, and a Historian. Data from sensors is available to the SCADA system and recorded by the Historian for subsequent analysis.

2.2.2.1 Layer 1 (L1)– Plant Control Network

Layer 1 refers to the communication among process PLCs. A star network topology used in level 1. This includes a SCADA workstation, HMI and a Historian.

2.2.2.2 Layer 0 (L0) – Process

Layer 0 refers to the communication between sensors, actuators and the PLCs. It is implemented by "Device Level Ring" which also includes a Remote IO (RIO) device. The RIO, and not a PLC, is connected to the physical sensors and actuators, with monitoring and control information being sent across the Distributed Logical Router (DLR). The ring topology allows active PLC controller to serve as "Ring Supervisor" and can tolerate single-node failure.

SWaT is equipped with Allen Bradley ControlLogix PLCs. Therefore, some of the attacks described below required consideration of the protocols used by these components, e.g., EtherNet/IP for Allen Bradley PLCs.

2.3 Attack Phase

From 27 August to 30 August, each Red Team was given 4-hours to demonstrate their attacks and achieve pre-determined goals (see Section 3 for details on scoring). The 4-hours included equipment installation, reconnaissance, designing and launching attacks, interactions with judges, and taking breaks.

Red Teams launched their attacks from inside the SWaT control room; only the active Red Team, accompanied by judges, iTrust staff and selected observers were allowed in the SWaT control room. Concurrently, an array of detection mechanisms was deployed in SWaT by the Blue Teams prior to starting the Exercise. These mechanisms monitored and attempted to detect and report, but not prevent, the ongoing attacks. The Blue Teams were in a separate room from the Red Teams to monitor and report the attacks to iTrust observers. The SWaT testbed was out of bounds to both the Red and Blue Teams. In addition, the SWaT Control Room was out of bounds to the Blue Teams.

Red Teams could interact with and attack most of the system except for the following.

a) physical access to SWaT and ZCC was not allowed;
b) server in the control room could not be attacked through physical means;
c) 10.10.0.0/16 and 1.2.222.0/24 could not be attacked;
d) the historian could not be compromised directly though manipulation of data sent to the historian was allowed; and
e) attacks could not affect the overall setup on a scale that affects beyond the provided testbed boundary (e.g. trigger university-wide fire alarm or similar).

3 Scoring

The scoring of the attacks was based on a point system. The equation below defines how points for an attack were awarded:

$$s = p * \left[\left(g_{p1} * c_{p1} + \cdots + g_{pn} * c_{pn}\right) + \left(g_{s1} * c_{s1} + \cdots + g_{sn} * c_{sn}\right)\right] + b$$

where p is the point awarded for entry modifier, g the points awarded based on whether specific physical process (g_p) or sensor data goals (g_s) can be manipulated, c is control modifier that is awarded based on the extent of control the attacker has in manipulating the physical process (c_p) or sensor data goals (c_s), and b denotes bonus points awarded for novel attacks outside of the physical process and sensor data goals listed.

3.1 Point of Entry Modifiers, p

a) 1: Entering via ZCC to launch an attack on SWaT
b) 0.8: Launching attack directly from SWaT

3.2 Goals $[g]$

3.2.1 Physical Process Goals: Control over physical actuator or process $\left[g_p\right]$

a) 100 points: Motorised Valves (open/close/transitioning/intermediate)
b) 130 points: Water Pumps (on/off)
c) 145 points: Pressure
d) 160 points: Water Tank Level (true water amount, not sensor reading)
e) 180 points: Chemical dosing

3.2.2 Sensor Data Goals: Demonstrate control over sensor readings at different components $[g_s]$

a) 100 points: Historian values
b) 130 points: HMI/SCADA values
c) 160 points: PLC values
d) 200 points: Remote I/O values

3.3 Control modifiers $[c]$

The control modifier determines the amount of control precision the attacker has during the execution of their attacks. As a guideline, the modifier is:

a) 0.5 if the team can only randomly influence the process (value and time)
b) 1.0 if the team can precisely influence the process or sensor value to a target value chosen by the judges

4 Description of Attacks Launched by Red Teams

4.1 iTrust Red Team, Singapore

The iTrust red teams consist of security researchers from iTrust and a PhD student candidate from Information Systems Technology and Design (ISTD) pillar.

4.1.1 Precise Control of Motorised Valve

Objective : Take control of MV201
Attack method : Man-in-the-middle (MITM) in level 1
Tools *:* netfilterqueue, wireshark, *bridge utils*
Description *:* The attackers attached their laptop to the switch and bridge the network. The attack was done by changing the value of MV201 that was transmitting from PLC1 to PLC2. The value was changed to open the valve.

4.1.2 Precise Control of Water Pumps

Objective : Take control of permeate pump (P602)
Attack method : Python Script
Tools *:* Python Script (pycomm)
Description *:* The intention of the attack was to open P601 so that more water would be added into the Raw Water tank causing overflow. The attackers launched a python script where to set the P601 to manual mode and the status of the pump to on.

4.1.3 Precise Chemical Dosing

Objective : Manipulate $NaHSO_3$ Pumps (P403 & P404) so that Hardness Analyser (AIT-401) remains high.
Attack method : Python Script
Tools *:* Python Script (pycomm)
Description *:* The intention of the attack is to switch on RO Transfer pump (P401) and close P403 & P404 so that no chemicals will be added to the water. The attacker launches a python script where these pumps are manipulated, by first setting the pump to manual mode and then set P401 to ON and P403 & P404 to OFF.

4.1.4 Manipulation of UF feed tank and Pump Values

Objective : To give contradicting values to PLC and the physical process
Attack method : exploits, Man-in-the-middle (MITM) in level 0
Tools : *Metasploit (eternal blue), netfilterqueue, wireshark, bridge utils*
Description : The intention of the attack is to trick the PLC and the SCADA workstation that there was no attack on P301 when in actual the UF feed pump (P301) was switched on. The attackers first changed the PLC code of stage 3 of the pump so that it did not produce an error. This done by eternal blue attack to gain access to the workstation, an eternal blue attack was launched. An admin account was created to login and then the PLC code was changed. The attackers then attached the laptop between the PLC and the RIO. The attackers spoofed the actual values of the tank and pump to the PLC as well as the command of the pump to turn on to the RIO.

4.1.5 Precise Control of PLC values

Objective: Manipulate Raw Water pump by spoofing UF feed sensor tank values (LIT301)
Attack method : Man-in-the-middle (MITM) in level 1
Tools : netfilterqueue, wireshark, *bridge utils*
Description : The attackers attached his laptop to the switch and bridge the network. The attack was done by changing the value of LIT301 that was going from PLC3 to PLC1. The change in value of LIT301.Pv to 700 resulted in P101 being switched on.

4.1.6 Manipulation of HMI Values

Objective : Remotely access SWat Network
Attack method : Bridging network interface
Tools : *Python script (Pycomm)*
Description : The attackers discovered that the internet access and the SWaT network were on the same interface. The IPs for internet and SWaT were 10.0.X.X and 192.X.X.X respectively. The attackers added both networks as sub-interfaces to access both. The attackers then forwarded port 22 to the internet and to establish a command and control server running in that network. This allowed the client to control the sensor and actuator values from a remote device through internet. A Pycomm script was then run to spoof the values of LIT101.

5 Description of Defence Teams

5.1 iTrust Anomaly Detection Mechanisms (ADMs)

5.1.1 Distributed Attack Detection (DAD)

5.1.1.1 Background

Distributed Attack Detection (DAD) is an attack detection system developed in-house by a team of researchers in iTrust. DAD is a product in development with its patent filed. Through the course of its development, DAD was iteratively improved through extensive experimentation in SWaT.

5.1.1.2 Technology Description

DAD can be considered as a host-based intrusion detection system (HIDS). Specifically, it collects data on the various sensor measurements of processes such as water pH, water level and flow indicator of the plant, for

analysis and process anomaly detection. By using measurements from 52 sensors in SWaT, it can detect single-stage multipoint and multi-stage multi-point cyber-attacks in a distributed control system.

DAD is novel because it uses "security by design" for many basic and advanced attacker models. Based on the laws of physics, it directly verifies the interactions among process variables of the plant within the distributed PLCs to check for abnormal behaviour. Process variables are time-dependent and interrelated within the entire plant process. Hence, their values are constrained by the relationship they have with the other process variables, as governed by the fundamental laws of physics and/or chemistry. The relationships among these constrained variables lead to process invariants and forms the backbone of DAD's rule-based algorithms.

The invariants are embedded in the PLCs as well as special hardware devices known as intelligent checkers (ICs) with wired interfaces to sensors and actuators. The invariants are checked constantly to ensure the underlying processes are behaving as intended. Violation of an invariant is indicative of divergence of the process from its internal behaviour

Figure 10 below shows an instance of the DAD's interface for which an alarm for invariant P1_SD5 was triggered as it was detected as being violated. In this example, the physical rule that is violated is part of the encoded control logic in SWaT, whereby the Raw Water pumps P101 and P102 should be turned on when the Ultrafiltration Feed Water Tank Level (LIT301) downstream is low.

Figure 30: DAD Alarm Screenshot

5.1.2 Argus

5.1.2.1 Background

Argus is a variant of DAD. Essentially, it has the same detection capability as DAD but relies on data from the Historian instead of from the PLC. In addition, Argus can be deployed on a separate server for an added defence layer towards orthogonal defence.

5.1.2.2 Technology Description

Argus leverages on the same principle of process invariants and uses the same algorithms employed in DAD. The difference is that Argus takes in data from Historian of a CPS. In this manner, Argus is still able to detect

process abnormalities when the PLC has been compromised. Argus is useful in operational legacy systems since legacy systems may not be able to support the deployment requirements for DAD.

5.1.3 GARX

5.1.3.1 Background

GARX is a defence mechanism based on the refinement of auto-regression and cumulative sum techniques. It detects anomaly for continuous systems in SWaT.

5.1.3.2 Technology Description

GARX use sauto- regression. In one step, it can predict the change in the endogenous variable y (whose value is determined by other variables in the system as the endogenous variable) based on a single delay for each exogenous variable x (whose value is determined by variables *outside* the system.). The value of the fitness function is between zero and one. A high value, for instance 0.98, indicates that the invariant model fits well with the observed data. When a change is detected the value is set to the sum variable that is added cumulatively. If an alarm is raised, its due to exceeding of threshold of the strategy used.

5.1.4 Multi-layer Perceptron (MLP) Neural Network Based Anomaly Detector

5.1.4.1 Background:

Multi-layer perceptron (MLP) neural network-based anomaly detector was developed for the real time detection of process anomalies in SWaT.

5.1.4.2 Technical description

MLP-based anomaly detector is an unsupervised HIDS that relies on sensor measurements for anomaly detection. Values of these state variables are time-dependent during the entire operation of SWaT. Thus, by treating the measurements as a time series prediction problem their temporal dependences are effectively captured by the MLP model for prediction. During the training process, several parameters (number of hidden units, hidden layers, learning rate etc.) of MLP are finetuned using the data collected from normal operation of plant to achieve minimal prediction error. Further, a window based cumulative sum and percentile approach is used to detect abnormal deviations between the observed and predicted sensor values for the identification of anomalies with minimal false alarms.

5.1.5 NoisePrint

5.1.5.1 Background

NoisePrint is a scheme proposed to detect data integrity attacks on sensors in plant. It uses a fingerprinting approach based on sensor and process noise.

5.1.5.2 Technical description

NoisePrint combines the fingerprints for sensor and process noise that are created during the normal operation of the system. As an example, under sensor spoofing attack, this noise pattern deviates from the fingerprinted pattern enabling it to detect attacks. To extract the noise (difference between expected and observed value) a representative model of the system is derived. A Kalman filter is used for the purpose of state estimation. By subtracting the state estimates from the measured system states, a residual vector is obtained. It is observed that in steady state the residual vector is a function of process and sensor noise. A set

of time domain and frequency domain features is extracted from the residual vector. Feature set is provided to a machine learning algorithm to identify the sensor and process.

5.1.6 HybMonitor

5.1.6.1 Background

HybMonitor is a detection method based on a novel modelling framework. It uses the model of the system under analysis to predict future behaviours. It can detect behaviours that diverge from the expected.

5.1.6.2 *Technology Description*

HybMonitor tool is a black-box modelling approach to detect cyber-attacks in Cyber-Physical systems. It relies on two different tools: HybModeller and HybMonitor. HybModeller uses historical data (data from historian) and creates a model of the normal behaviour of the system. The second component (HybMonitor) uses system's models and predicts 'normal' behaviour of the system under test. It reads the actual state of the system, identifies the operational mode and predicts sensor readings. HybMonitor can predict state transitions in a controller based on prior knowledge.

5.2 Commercial Products

5.2.1.1 Background

Four commercial vendors, referred as A, B, C and D, participated in CISS as Blue Teams.

Product A's detection mechanism is based on IP-based communications that pass through the system.
Product B – not revealed
Product C is designed to detect all the three types of attacks (physical process anomalies, sensor data anomalies and IT anomalies) via detecting their impact on the network traffic.
Product D detects all the three types of attack as well. It detects the attacks on the physical and I/O layer and complements the detections on the network layer. On the IT layer, it can detect and co-relate with processes to ensure that L0 to L3 are covered.

6 Evaluation of Defence Mechanisms

The defence mechanisms mentioned above were each evaluated based on (a) detection score and (b) breadth of defence as presented in this section.

6.1 Detection Score and Breadth of Defence

The **detection score** of each defence mechanism was computed as a metric of the actions and attacks by the Red Teams. Actions and attacks that were not successful were also accounted for as these actions and attacks would create some noise and affect the system. Detecting such actions and attacks would be considered a successful detection of an anomaly. The false positive rates of iTrust Technologies are also presented; however, the false positives for commercial teams are not included as such data was not provided to iTrust.

The Blue Teams were asked to cross-reference the time-stamped attack logs (supplied by iTrust) to their detection logs so as to validate detection. The Blue Teams were also asked to provide evidence for each attack they claim was detected. The criteria - if an alarm attributable to the anomaly is raised within 5 minutes of the time at which the attack was launched, the anomaly was considered detected. Hence, only immediate and

apparent detections were credited as successful detections. Such examples of detections are described in the following subsections 6.1.1 - 6.1.4.

The **breadth of defence** was analysed by determining how well the detection mechanisms detected attacks over ten different categories, namely, the five Physical Process Goals and four Sensor Data Goals (described in Section 3.1,) and IT based actions and attacks.

6.2 Summary of Results

Tables 1 and 2 below tabulate percentages of attacks detected by each detection mechanism for each attack category.

Table 1: Percentages of Anomalies Detected by Products A to D

Category of Anomalies	No. of Anomalies Recorded	Percentage of Anomalies Detected			
		Product A	Product B	Product C	Product D
Physical Process Goals					
a) Motorised Valves	13	62%	38%	38%	85%
b) Water Pumps	21	71%	62%	29%	76%
c) Pressure	3	67%	33	0%	67
d) Water Tank Level	14	86%	86%	43%	93%
e) Chemical Dosing	4	25%	0%	0%	100%
Sensor Data Goals					
a) Historian values	1	100%	0%	100%	100%
b) HMI/SCADA values	7	0%	0%	43%	71%
c) PLC values	24	50%	50%	71%	92%
d) Remote I/O values	3	0%	0%	67%	67%
IT Anomalies	54	35%	43%	44%	63%
OT Anomalies	90	57%	48%	45%	85%
Total Anomalies	144	49%	46%	45%	76%
False Positive		Data not available			

Table 2: Percentages of Anomalies Detected by iTrust ADMs

Category of Anomalies	No. of Anomalies Recorded	Percentage of Anomalies Detected					
		Argus	DAD	GARX	MLP	HybMonitor	NoisePrint
Physical Process Goals							
a) Motorised Valves	13	0%	23%	0%	0%	46%	0%
b) Water Pumps	21	0%	24%	19%	0%	24%	5%
c) Pressure	3	0%	33%	0%	0%	33%	0%
d) Water Tank Level	14	21%	57%	29%	43%	14%	43%
e) Chemical Dosing	4	0%	25%	0%	0%	0%	0%

Sensor Data Goals							
a) Historian values	1	0%	0%	100%	0%	0%	0%
b) HMI/SCADA values	7	0%	43%	0%	29%	0%	0%
c) PLC values	24	17%	38%	17%	21%	0%	0%
d) Remote I/O values	3	33%	100%	33%	0%	0%	33%
IT Anomalies	54	2%	2%	2%	0%	0%	0%
OT Anomalies	90	8%	36%	10%	14%	15%	9%
Total Anomalies	144	6%	23%	10%	9%	10%	6%
False Positive		50%	52%	17%	57%	88%	56%

Results of Blue Teams' performance are summarised in Figure 11.

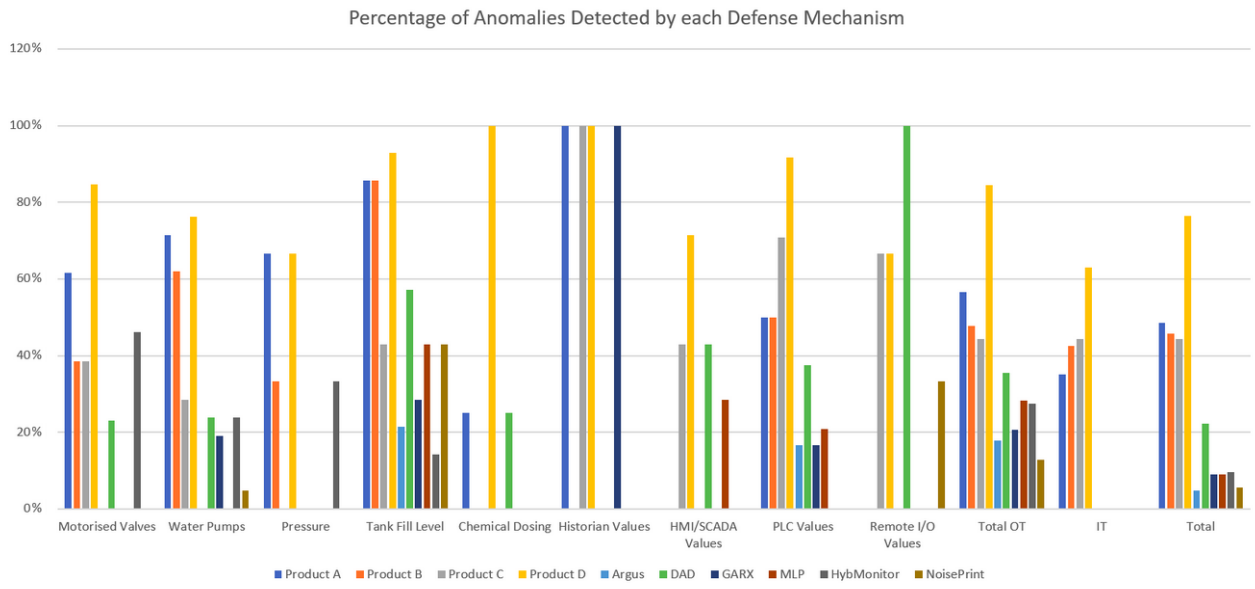


Figure 41: Percentages of Anomalies Detected by each commercial product and iTrust ADMs

Table 2 shows iTrust ADMs' performance irrespective of whether they were designed to detect those types of attacks. Table 3 reproduces Table 2 and excludes detection rates (by replacing them with "X") for attacks that the detection mechanism was not designed for e.g., sensor data and IT anomalies.

Table 3: Percentages of Anomalies Detected by iTrust ADMs based on design

Category of Anomalies	No. of Anomalies Recorded	Percentage of Anomalies Detected					
		Argus	DAD	GARX	MLP	HybMonitor	NoisePrint
Physical Process Goals							
a) Motorised Valves	13	X	23%	X	X	46%	X
b) Water Pumps	21	X	24%	19%	X	24%	x
c) Pressure	3	X	33%	X	X	33%	X

d) Water Tank Level	14	21%	57%	29%	43%	14%	43%
e) Chemical Dosing	4	X	X	X	X	X	X
Sensor Data Goals							
a) Historian values	1	X	X	X	X	X	X
b) HMI/SCADA values	7	X	X	X	X	X	X
c) PLC values	24	X	X	X	X	X	X
d) Remote I/O values	3	X	X	X	X	X	X
IT Anomalies	54	X	X	X	X	X	X
Total OT Anomalies		21%	33%	23%	43%	27%	43%

6.2.1 Detection on Physical Process Anomalies

Physical process anomalies consist of (a) motorised valves, (b) water pumps, (c) pressure, (d) water tank level and (e) chemical dosing. In total, 54 physical process anomalies were recorded.

Across the commercial products, Product D outperformed the other products. It had a 100% detection rate for (e) and superior detection rates for attacks (a) to (d). Product C performed the poorest; it was unable to detect attacks (c) and (e).

Across iTrust ADMs, HybMonitor and DAD were the two top performers. HybMonitor outperformed DAD in attack (a) while DAD fared better for attacks (e). Both HybMonitor and DAD were tied (b) and (c). HybMonitor detect anomalies based on the control logic. If the physical process deviates from the control logic that was trained in, it will lead to the detection.

The number of anomalies recorded consists of both successful and unsuccessful attacks as well as attempted attacks. Based on table 3, when comparing iTrust ADMs with the commercial products based on the anomalies recorded (both successful and unsuccessful), Product D outperformed iTrust ADMs.

6.2.2 Detection on Sensor Data Anomalies

The detection of sensor data anomalies consists of a) historian values, b) HMI or SCADA values, c) PLC values and d) Remote I/O values. In total, 35 sensor data anomalies were recorded.

With respect to sensor data anomalies, Product D performed significantly better than all other technologies; Product C came in second best. Product D detected 30 out of 35 attacks i.e., about 85% of the attacks. Product C followed with 65% detection rate. Among iTrust ADMs, 'DAD' was the best performing defence with 43% detection rate. iTrust ADMs seemed to perform poorly because they were designed to detect process anomalies, and in most of the attacks even though the sensor data was affected, no process anomaly resulted and hence not detected.

6.2.3 Detection on IT Anomalies

The detection of IT anomalies consisted of various IT actions like port scanning and IT attacks like ARP spoofing and EternalBlue exploit. In total, 54 IT anomalies were recorded. In general, Products A to D performed much better in detecting IT Anomalies as compared to iTrust ADMs. This is due to Products A to D listening to the network packets on Level 1 and training their defence mechanism to create a baseline of communication

across devices. iTrust ADMs, however, performed poorly in this area as they monitor only measurements from the PLC and Historian.

6.2.4 Design of iTrust ADMs

iTrust ADMs are designed for only OT based attacks. Each ADM is designed differently for a different purpose and hence meant to detect specific types of attacks. In Table 3, the "X" denotes the attacks that the ADMs are not designed to detect. For example, GARX is designed based on the level sensors of the plant and thus will not be able to pick up attacks affecting chemical sensors. Similar, HybMonitor focuses on physical goals instead of the sensor goals. Also, as most of these ADMs are ongoing research, their detection capabilities were still works in progress at the time of CISS 2019. Table 3 shows that iTrust ADMs failed to detect a significant number of anomalies. Several possible reasons are given as follows:

Argus:
Argus detects sensor reading fluctuations when discrepancies are wider (5-10mm for Level sensors). Some attacks on level sensors (LIT) kept the level sensor fixed for just a few seconds to demonstrate the attack to the judges and then went back to normal. Such attacks remained undetected by Argus. The attacks that were detected were those that cause a significant amount OT disruption. In addition, Argus faced some technical issues such as logging which displaced timestamps of detection when run for extended periods. The logging system was not fully automated during CISS but has been improved since then. Argus was also not in operation on Day 2 afternoon session resulting in no detections during that session.

DAD:
DAD did not perform to its best due to miscommunication with the author of the implementation of DAD during the event. It was later discovered that the implementation of the technology was the beta version of DAD. The false positives of DAD was due to the invariants being incorrectly coded.

GARX:
GARX only detects specific attacks that relate to continuous valued state variables, e.g., LITs. Hence, attacks that do not trigger such variables were not detected.

MLP:
The MLP based anomaly detector was designed to monitor and detect real-time anomalies on water level sensors. The reason why few attacks were not detected was due to their stealthy nature. The attacker injected the false measurement in such a way that it was close to the normal operation of SWaT tested. Thus, in order to detect such attacks, the higher-order dependencies like inflow and outflow rate, status of the pump and motorized valve need to be considered.

HybMonitor:
HybMonitor showed the highest amount of false positives among the iTrust ADMs. This was due to error propagation of the model. The author plans to deploy feedback control to prevent the cascading effect and reduce the number of false positives. Additionally, HybMonitor identifies attacks via clustering of multiple anomalies during a period, with a high number of false positives, new attacks can be ignored because the tool misclassifies them as previous events. An improved method for attack identification is in design.

NoisePrint:

NoisePrint was designed using data from the SWaT historian. For the earlier published results of NoisePrint [AsiaCCS-2018] data was taken from the SWaT historian at a sampling rate of one sample per second. During the CISS event NoisePrint received the samples from Clone Historian[1] where sample frequency was between 1-3 seconds depending on the number of requests to Clone Historian. The noise pattern being sensitive to sampling rate led to unexpected performance. Secondly, the recent data used to create models was collected for the states when the plant was operational, the author realised during the CISS event when a lot of false alarms were observed when plant was not active. In future, the author can obtain the data from the plant for all states (active and idle) using the Clone Historian but not the historian.

6.2.5 Combined iTrust ADMs

Table 4 shows how iTrust ADMs perform when they are combined as a single ADM, since each ADMs is designed to detect different types of anomalies. Anomalies in network traffic were omitted as iTrust ADMs are not designed to detect these. iTrust technologies performed reasonably well in successful detection of attacks to what the technology is designed for; they obtained 100% a) and d) in physical process goals.

Table 4: Evaluation of an integrated complementary set of iTrust Technologies

Category of Anomalies	No. of Anomalies Recorded (Successful & unsuccessful)	ITrust Technology Anomaly Detection Percentage	No of Anomalies Recorded (Successful)	iTrust Technology Success OT Attacks Percentage
Physical Process Goals				
a) Motorised Valves	13	29%	5	100%
b) Water Pumps	21	52%	11	82%
c) Pressure	3	33%	2	50%
d) Water Tank Level	14	57%	5	100%
e) Chemical Dosing	4	-	1	-
Sensor Data Goals				
a) Historian values	1	-	1	-
b) HMI/SCADA values	7	-	6	-
c) PLC values	24	-	11	-
d) Remote I/O values	3	-	3	-
IT Anomalies	54	-	38	-
OT Anomalies	90	29%	46	41%
Total based on technology	51	43%	23	83%

7 Conclusion

Seven Red Teams and five Blue Teams (4 commercial products and an iTrust Blue Team featuring six ADMs). The Red Teams were awarded points based on the type of attacks launched, point of entry, physical process and sensor data goals and the level of precision the attacker had during the execution. The Blue Teams were evaluated based on the number of anomalies detected arising from successful and unsuccessful attacks. The Blue Teams were scored using the submitted evidence (of attack detection), based on the attack logs provided

[1] Clone Historian acts as a Client for the primary Historian and a Server for Blue Teams and allows multiple requests to retrieve data from the SWaT Historian, thereby reducing the sampling load on the Historian

by iTrust. Some evidence were not conclusive and there is a possibility that they might not have been able to distinguish between an anomaly or noise without those attack logs. Nevertheless, CISS has exposed these defence mechanisms to skilled attackers and the environment to be tested on. It enabled the Blue Teams to show case / perform empirical tests on their defence mechanisms and the Red Teams to have a unique chance to attack a realistic process plant.