The Fourth International

CISS

Critical Infrastructure Security Showdown - Online 2020

Technical Report

Sponsors:

National Research Foundation, Singapore

Ministry of Defence, Singapore

Date:

July 27 - Aug 7, 2020

Report by:

Ken CHIN, Beebi Siti Salimah Binte LIYAKKATHALI, Francisco FURTADO, Ivan LEE, THUR You Fu, Yoga Kashenen s/o YOGAINDRAN

Organiser:

iTrust, Centre for Research in Cyber Security

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1 The Critical Infrastructure Security Showdown– Online 2020 (CISS2020-OL), conducted over two weeks from July 27 - Aug 7, 2020 at the Singapore University of Technology and Design (SUTD), was the fourth run of iTrust's technology assessment exercise. Owing to the pandemic, CISS was moved to an entirely online platform. Doing so allowed participants – red and blue teams – to still enjoy the same level of access and experience to the exercise platform as if they were physically onsite.

1.2 CISS2020-OL was sponsored by Singapore Government agencies, the Ministry of Defence and the National Research Foundation.

2. Objectives

2.1 CISS2020-OL's key objectives are to: (a) validate and assess the effectiveness of technologies developed by researchers associated with iTrust; (b) develop capabilities for defending critical infrastructure under emergency situations such as cyber-attacks; and (c) understand composite Tactics, Techniques and Procedures (TTPs) for enhanced Operation Security (OpSec).

2.2 From the perspective of participating red teams, CISS2020-OL provided hard-to-get insight and hands-on experience in understanding the approaches required for compromising critical infrastructure.

3. Phases in CISS2020-OL

CISS2020-OL consisted of the following time-sequenced phases:

Phase I [May 4 - 29, 2020]	:	Participant selection (red & blue teams, observers)
Phase II [June 22 - July 3, 2020]	:	Participant familiarisation (red & blue teams)
Phase II-A [June 22]		Blue team briefing
Phase II-B [June 29]		Red team briefing
Phase II-C		Judge briefing
Phase III [July 6 - 16, 2020]	:	Target system selection (red teams)
Phase IV [July 27 - Aug 7, 2020]	:	CyberFire (red & blue teams, observers)

Phase V [Q3 – Q4, 2020] : Data analysis and reporting

Throughout the document there will be several mentions of the tools deployed by iTrust to manage the entire exercise. Readers are encouraged to familiarise themselves with these terms by referring to paragraph 10 of this report.

3.1. Phase I: Participant selection

3.1.1. Participation in CISS2020-OL was by invitation only. Participants were classified into:

a) Red teams (up to 6 members per team):
   - 18 local and international teams from government organisations, private sectors and academia, including two from iTrust.
   - Participating teams were from Finland, France Netherlands, Poland, Singapore, South Korea and United States of America.

b) Blue teams (no limit on the number of members):
   - 12 teams from private sector and academia, including 6 teams from iTrust

c) Remote observers: Singapore Government agencies and their invitees.

3.2. Phase II: Participant familiarisation

3.2.1. All red and blue teams were given an online tour of the Secure Water Treatment (SWaT) testbed. Briefing and Q&A sessions were also organised for red and blue teams.

3.2.2. In addition, the red and blue teams were provided documented information on SWaT, the digital twin, digital twin player, and various anomaly detection and plant safety technologies that would be deployed during the exercise.

3.2.3. Blue teams that needed to install their hardware and learn the normal behaviour of SWaT were given sufficient time to do so. They were required to adhere the following:

- The installations (hardware and software) do not disturb the regular plant operation and interfere with existing iTrust technologies;

- Make its own arrangements for the data generated by its hardware to be piped to their computers outside of the SWaT during the exercise;
- The installations respond as if in a real-time environment;
- The installations be completely removed post-exercise and restore SWaT to its original condition. The blue team would bear any cost for damages arising from the installation and/or teardown of the upgrades; and
- Disallowed to prevent, halt or thwart any attacks launched by the red teams.

3.2.4. Blue teams' systems were connected to iTrust's TAP switch to receive pcap data from Zycron Cyber City and SWaT (see Figure 4). For details on attack detection and reporting by blue teams, refer to paragraph 3.4.4.2.

3.3. Phase III: Target system selection

3.3.1. During this Phase III, each red team was provided with 7 instances of data collected from SWaT and its digital twin, referred to as Target 1, Target 2…Target 7. A higher score was given if the red team successfully selected SWaT as the target (see paragraph 3.4.4 for details on scoring.)

3.3.2. Red teams could choose to analyse the dataset in one of the following options:

a) Option 1: Use the 2-hour slot to connect to Cloud VM to view the datasets and "play" the datasets view a player; or

b) Option 2: Download the datasets 48 hours prior to their 2-hour slots and analyse them.

3.3.3. Details of both options are as follows:

<u>Option 1: Use 2-hour slot to connect to Cloud VM</u>

3.3.4. Red teams were provided unique credentials to connect to Cloud VM 30 minutes before their target selection timeslot. OT data captured by the historian by each target system i.e., Target 1, Target 2…Target 7 from SWaT and the digital twin historians were available on the Cloud VM and could be viewed through PEPPR-PV and PEPPR-PP. Figure 1 on the next page captures the interactions between a red team and the targets. Note that ZCC (see para 3.4.1) was not available during this phase.

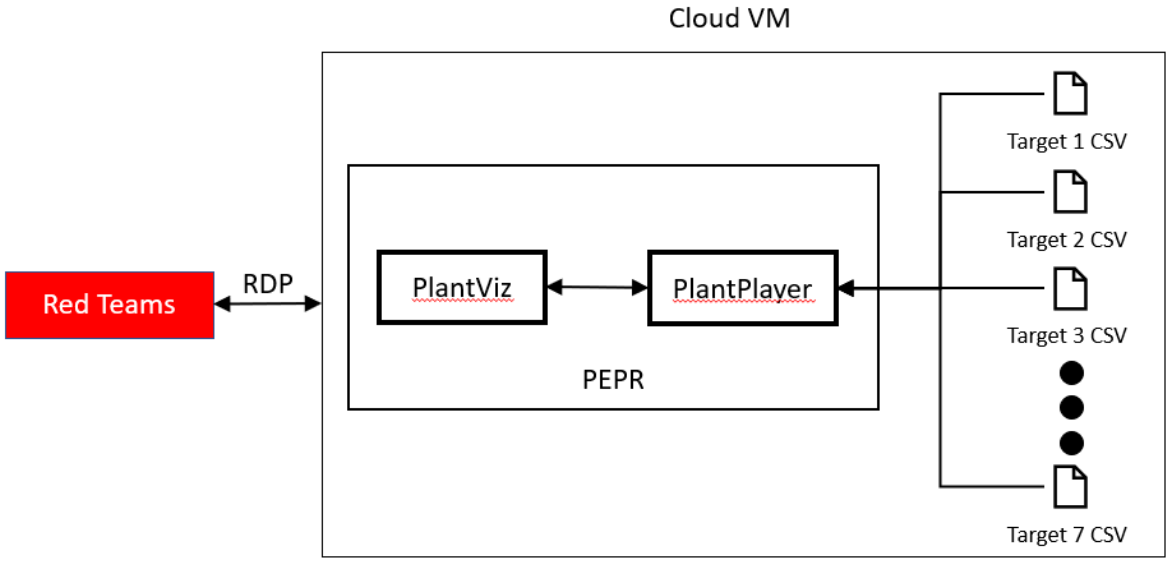


Figure 1: Interactions between red team & CISS2020-OL system & tools in target selection

3.3.5. Red teams were provided unique credentials to connect to Cloud VM 30 minutes before their target selection timeslot. OT data captured by the historian by each target system i.e., Target 1, Target 2…Target 7 from SWaT and the digital twin historians were available on the Cloud VM and could be viewed through PEPPR-PV and PEPPR-PP.

3.3.6. Each red team was asked to make known their target system selection to iTrust within 2 hours from the end of their target selection slot.

<u>Option 2: Download the datasets 48 hours prior to their 2-hour slots</u>

3.3.7. If the red team selected this option, the datasets were available for download by the red team 48 hours before its target selection slot (e.g. if the slot it chose was Wednesday, 4pm (GMT+8) then the link to download the dataset would be sent to it on Monday, 4pm (GMT+8). The red team was given 48 hours to make known its selection to iTrust.

3.3.8. Regardless of its selection (whether it chose SWaT or digital twin) during this phase, all red teams were given the full four hours to launch attacks on SWaT.

3.4. Phase IV: CyberFire activities

The CyberFire activities were spread over 16 CFM (Table 1). The duration of each CFM slot was 4 hours, from 9am to 1pm or from 2pm to 6pm, GMT+8, with a one-hour break in between for system reset.

Table 1: CISS2020-OL Schedule for red teams

Week 1			Week 2	
Date	**CFM slot**		**Date**	**CFM slot**
Mon July 27	1 (AM)		Mon Aug 3	9 (AM)
	SR			SR
	2 (PM)			10 (PM)
Tue July 28	3 (AM)		Tue Aug 4	11 (AM)
	SR			SR
	4 (PM)			12 (PM)
Wed July 29	5 (AM)		Wed Aug 5	13 (AM)
	SR			SR
	6 (PM)			14 (PM)
Thu July 30	7 (AM)		Thu Aug 6	15 (AM)
	SR			SR
	8 (PM)			16 (PM)
Fri July 31	No activity; Public holiday		Fri Aug 7	17 (AM)
				SR
				18 (PM)

CFM: CyberFire Module; red teams attack a target system; SR: System reset (1 hour)
AM slot: 0900 – 1300; PM slot: 1400 – 1800, GMT +8

3.4.1. Attack platform

For added realism, all red teams were required to attack SWaT by first entering the network via the ZyCron Cyber City (ZCC; Figure 2); they would land in ZCC's corporate network through a VPN connection. ZCC is a full-fledged virtual organisation comprising Information Technology (e.g., e-mail server, file server, printer server, CCTV, honeypot and intranet) and Operational Technology (processes similar to those in SWaT). To make these entities "alive," various types of network traffic were also crafted and included in ZCC. As an IT environment, ZCC was not set up with best practices i.e., it was intentionally built with minimum security features and contained vulnerabilities for red teams to explore and exploit. There was no internet access within the ZCC.

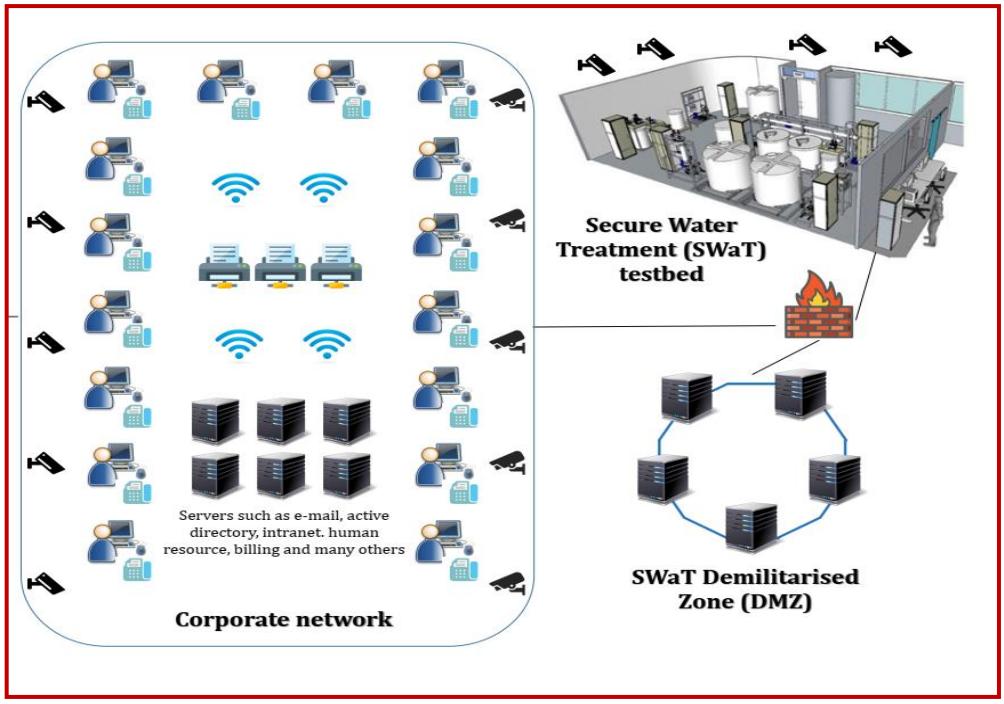


Figure 2: High-level Architecture of ZyCron Cyber City

3.4.2. Launching attacks

3.4.2.1. Active stage

During a CFM the active red team was asked to demonstrate its attacks and achieve the pre-determined goals (see para 3.4.4 for details on scoring). The CFM duration included, but was not limited to: reconnaissance, designing and launching attacks, interactions with judges (e.g., for Attack Logging; see Figure 3) and taking breaks.

3.4.2.2. Hunting stage

All red teams had to enter SWaT via the ZCC to launch attacks. Failure to do so and to identify the pre-selected target system led to a lower score. If, during its CFM slot, attempts to penetrate into SWaT network through ZCC corporate network were unsuccessful after 30 mins, the red team could request to extend to up to 60 mins. If the attempt was still unsuccessful, iTrust assisted to port the red team over to SWaT.

ZCC is built with typical enterprise vulnerabilities that exist in many organisations. The red team had to "hunt" for these vulnerabilities and compromise them before using them to "hop" deeper into the network and eventually locate SWaT/digital twin in the OT network.

3.4.2.3. Attack launch stage

Active teams had four hours to design and launch attacks on SWaT (see Figure 3).

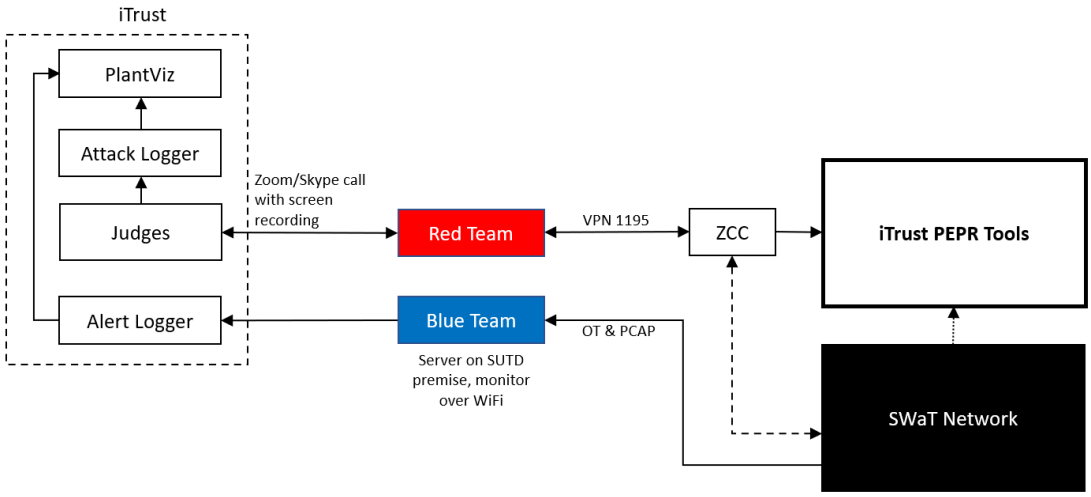


Figure 3: Interactions between red/blue teams & CISS2020-OL systems & tools in CyberFire

Prior to launching attacks, the active red team were required to do the following throughout its CFM:

a. Share with iTrust the "live" screen of the computer that is used to launch the attack via an online communication tool (e.g. Skype)[1];
b. Allow iTrust to video record the screen; and
c. Inform judges (1) the intention of the attack; (2) the targeted component(s); and (3) the launch procedure.

The duration of an attack was determined in real time by iTrust's cyber security technology engineers stationed physically at SWaT. Attacks that took a long time, e.g., 30 minutes, to have a noticeable impact on the plant could be halted by the judges before the impact was visible.

3.4.3. Attack monitoring

Any anomaly resulting from the attack, or otherwise (i.e., a false alarm), and reported by one or more iTrust detectors, was visible only to the organisers, observers and judges and not to the red or blue teams.

[1] For iTrust's post-event analysis and report writing purposes; recordings are not shared with anyone or made public without written permission by the red team

3.4.4. Scoring of red teams

The performance of each red team was assessed in real time by a team of judges comprising cyber security experts and engineers working in the critical infrastructure domain. Only single attacks, in series (one starts only after the previous one has ended), were allowed. Judges scored each red team based on criteria such as complexity of the attacks launched and success of the attack in resulting in an anomaly in at least one of the plant state variables. The total score, *S*, for each attack launched is computed based on five factors *t, p,* a_t, a_{sd} *and b.* These are described in detail below.

$$Total\ score,\ S = t + p * (a_{t1}a_{sd1} + a_{t2}\ a_{sd1} \ldots\ a_{tn}\ a_{sdn}) + b$$

where:

- *t* = target selection modifier
    - *Selected SWaT (t = 150) or one of the digital twins (t = 0) during target selection*
- *p* = point of entry modifier
    - *All red teams must attack SWaT by first entering the network via the ZCC (para 3.4.2.); p = 1*
    - *If attempts to enter ZCC are unsuccessful after 30 mins (request to extend to up to 60 will be considered), the team may proceed to attack SWaT or the digital twin (whichever was selected as the target in the selection phase) directly; p = 0.75*
- a_t = an **attack target is** a physical component or parameter in the plant on which the red team wants to launch the attack. An attack target differs from the **attack intention** which is defined as the intended impact as a result of the attack on the target. For example, to cause a water tank to overflow (attack intention), an attacker may choose to launch an attack on a valve (attack target) by setting it to the CLOSED condition long enough, without getting detected, so that a continuous flow of water into the tank is maintained.
- The 12 attack targets[2], and their corresponding points in parentheses, if an attack

[2] Activities or actions that would interfere, obstruct or disturb Participants, iTrust and running of the Exercise were strictly prohibited. In addition, the following were unavailable for attack:
- Hypervisors
- 10.10.0.0/16

is successful, in SWaT are:

- Conductivity meter (300)
- Flowmeter (200)
- Historian[4] (100)
- Water level meter (200)
- Oxidation Reduction Potential Meter (300)
- pH meter (300)
- PLCs (100)
- Pressure meter (200)
- Pumps (200)
- SCADA (100)
- Network switches (100)
- Valves (200)

- a_{sd} = attack success and detection modifier: whether an attack results in an anomaly, and whether the anomaly/attack is detected by any of the installed detectors.
  - $a_{sd} = s * d$
  - *If the attack is successful, s = 1; else s = 0*
  - *d is calculated as:*

↓ d	s →	Attack results in an anomaly	Attack does not result in an anomaly
Anomaly/attack is observed*		0.7	-0.2
Anomaly/attack is not observed*		1	0

**through physical observations of the plant and SCADA screen by plant operator and judges*

- b = bonus points for novel attacks (such as the ability to disrupt the anomaly detectors), at the discretion of the judges

3.4.4.1. Attack detection

Throughout the event the blue teams monitored their systems remotely (Figure 4 next page). Post-event, blue teams were given pcap and OT data captured for analysis.

- 1.2.222.0/24
- 9.9.0.0/16
- Server rack: The server rack should not be attacked through physical layer
- Historian: Comprising historian not allowed
- General electric supply, fire alarm systems etc.

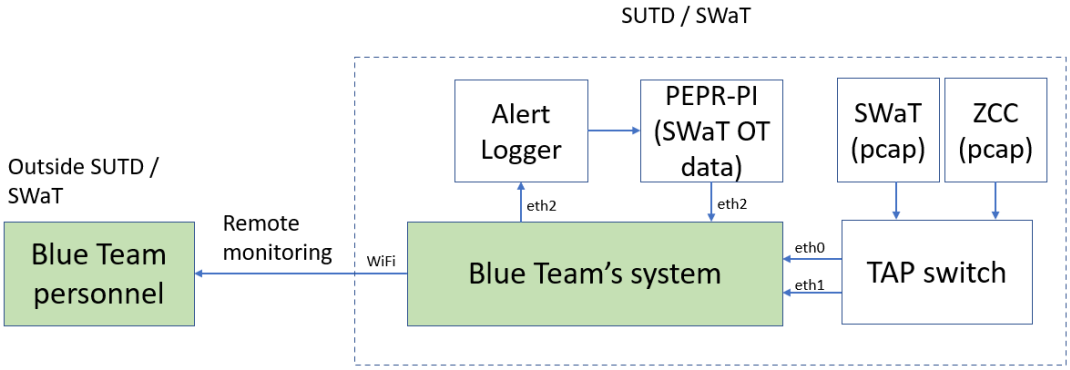


Figure 4: Blue teams remotely monitoring their systems' GUI

3.4.4.2. Reporting of alerts

The above assumption implies that any alert generated by the security system deployed by a blue team was reported *immediately* to the plant operator *automatically, not manually*. While each blue team was provided event data at the end of the event, they were not expected to conduct an analysis of an alert generated during the event. Blue teams were requested to report each alert immediately as if it were occurring in a live plant and being reported to the plant operator.

Reporting of alerts to iTrust by blue teams was done in one of the following two ways:

a) PEPPR-PV: this required the blue team to work with iTrust's developer to integrate with it, so that its detections/alerts could be sent to PEPPR-PV for automatic logging and visual alerts; or
b) Alert logger: a simple password-protected interface to manually log a time-stamped alert each time the blue team detected an attack.

3.5. Phase V: Data analysis and reporting

3.5.1 Data from each red team session were recorded and saved in the iTrust data library. These consisted of measurements from all sensors in SWaT as well as network packets saved into pcap files. Note that the recorded data contains data mutated by the red teams.

3.5.2 In the following sections, details regarding the types of attacks launched by red teams, description of the blue teams' mechanisms and the analysis of detections made by the blue teams are reported. The analysis will result in metrics such as the number and

types of attacks launched, success rate, detection rate (and false positives), and time taken to detect.

4. Description of Defence Teams

4.1. iTrust Anomaly Detection Mechanisms (ADMs)

4.1.1. Distributed Attack Detection (DAD)

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.

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 53 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 5: DAD Alarm Screenshot

4.1.2. HybMonitor

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.

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.

4.1.3. AI Crit

Background

AICrit's intelligence integrates the design knowledge and machine learning algorithms into one versatile solution for automated process monitoring and threat detection in the operational Industrial Control Systems (ICS).

Technology Description

AICrit for anomaly detection in ICS is a unified framework for real-time process monitoring with a goal to preserve the control behaviour integrity of the ICS. It precisely learns the normal spatio-temporal relationship among the set of highly correlated components through the application of machine learning algorithms (data-centric approach) and with a considerable amount of design knowledge (design-centric approach). The process involved in the design of the unsupervised detector presented here is of two-folds. One is modelling the normal behaviour of continuous-valued state variables (sensors) through the temporal dependencies to forecast their behaviour with minimal error. Second is modelling the higher-order and non-linear correlation among the discrete and continuous type state variables (cross-correlation among the sensors and actuators) during the normal plant operation. By combining these two, the functional dependencies of the sensors and actuators are monitored continuously, which increases the confidence in discovering and reporting a wide range of anomalies during the discrepancies in the expected and actual behaviour of ICS.

4.1.4. AEGIS

Background

Automatic Extensible Generic Invariant-based Security (AEGIS) is an attack detection tool that intends to augment the usability of DAD by automating the process of invariant creation. It comprises an algorithm designed to be generic and universal for various types of CPS, offering the option of plant-specific customisation for users.

Technology Description

The first step in the automation using AEGIS' algorithm hinges on the idea of reading the connections between the components of the plant from its CAD (Computer-aided Design) file or P&ID (Piping and Instrumentation Diagram). Based on encoded physics principles, the

algorithm then automatically generates the rules that the associated sensor-actuator sets must follow for the proper operation of the system. These rules, called invariants, are created using similar logic as followed by DAD.

When the tool is in operation, it keeps checking the incoming sensor and actuator readings to determine whether the actual system behaviour is in accordance with its expected performance. The violation of the invariants could be a sign of the presence of process anomalies, which could be occurring due to attacks.

The tool is modular in its architecture and allows plant operators to tune the generalised design parameters and device-specific constants to tailor the detector for their particular systems.

4.1.5. Attester

Background

The attestation tool is a mechanism, which specifically addresses the problem of attesting the integrity of the PLC code. The attestation techniques typically can be categorized into three types-software-based attestations, hardware-based attestation, and physical attestation. Since SWaT does not provide the hardware for hardware-based attestation, and access to the firmware, a practical remote attestation solution is used where mechanism only requires all sensor readings, actuator states, and variables concerning PLC state as input.

Technology Description

Firstly, the faithful offline copies of its PLC programs are written in python. This code will generate the corresponding actuator commands for the given sensor readings, actuator states, and variables. Based on the inputs from the real system, Attester tool can predict the state of the actuators. Then, the prediction states and the state in the future are checked and will raise alarm(s) if these two values are not consistent.

4.2. Commercial Products

Five commercial vendors, referred to as External Blue Teams (EBTs) 1, 2, 3, 4 and 5 participated in CISS as Blue Teams. Their identities are not revealed in this report.

5. Evaluation of Defence Mechanisms

The defence mechanisms were evaluated based on the total OT anomalies detected on the SWaT system at a particular Red Team session. IT anomalies were not considered as part of the evaluation. During the event, Blue Teams sent alarms live to the Alert logger when their defence mechanisms detected an anomaly. These alarms were then evaluated for True positive (TP) implying that the alarm is correct, False Positive (FP) when there is an alarm but no anomaly and False Negative (FN) for unable to detect an anomaly.

Each detector was scored accordingly to the red team sessions. The nomenclature for each red team session is DATE|SLOT. For example, Aug3AM would represent a session on the morning of the 3rd of August.

The 5 iTrust detectors were evaluated using the Aug3AM and Aug6AM session. These sessions were carried out by iTrust red team and contained numerous OT anomalies and attacks.

6.1 OT anomalies

For Aug3AM there were a total of 26 Anomalies; for Aug6AM there were 6. These anomalies are listed in Table 2 below.

Table 2: Anomalies occurred on Aug3AM

No.	Anomaly	Explanation
1	Anomaly in raw water inlet value and raw water outlet flow meter	Value of FIT101 was changed from 0 to 1.5 and MV101 from 1 to 2
2, 11, 17	Anomaly in ultra-filtration stage	Values of stage 3 was changed
3	Anomaly in NaHSo3 dosing pump	Value of P403 changed from 1 to 2; P401 and P501 started together; UV401 was OFF
4	Anomaly in raw water motorised valve	Value of MV101 was changed from 1->0->2->0->1->0->2->0->1->0->2>0->1.
5	Anomaly in speed of variable speed pump	Value of VSD pump was changed
6	Anomaly in ultraviolet dechlorinator	Value of FIT401 was changed to 0.49
7	Anomaly in UF feed water level meter	Value of LIT301 was changed to 780
8	Anomaly in HCl concentration	Value of P203 and P204 are changed to turn ON

9	Anomaly in RO permeate pressure	Value of PIT503 was changed
10	Anomaly in chemical dosing pH meter	Value of AIT202 value was changed
12	Anomaly in RO feed ORP meter	Value of AIT402 was changed
13, 26	Anomaly in de-chlorination stage	Values of stage 4 was changed
14	Anomaly in ultra- filtration	Value of LIT301 was changed; P101 was ON but LIT301 was not increasing.
15, 21	Anomaly in reverse osmosis stage	Values of stage 5 was changed
16	Anomaly in UF backwash valve	Value of FIT301 behaving abnormally
18	Anomaly in pre-treatment stage	All value of stage 2 pumps are 2
19	Anomaly in pre-treatment stage	Value of stage 2 pumps are reversed
20	Anomaly in dechlorination stage and valves of ultra- filtration stage	Error in Stage 3 Valves; Values of stage 4 was changed
22, 25	Anomaly in raw water level meter	Value of LIT101 was changed to 608
23	Anomaly in RO feed water level meter	Value of LIT301 was incorrect
24	Anomaly in RO feed water level meter and UF feed water level meter	LIT301 and LIT401 sensors are swapped

Table 3: Anomalies occurred in Aug6AM

No.	Anomaly	Explanation
1, 2	Anomaly in ultra-filtration stage	Values of stage 3 changed
3	Anomaly in raw water stage	P101/P102 is not running when LIT301 is low
4	Anomaly in level meter	Value of LIT101, LIT301, LIT401 was changed
5	Anomaly in UF feed water level meter	Value of LIT301 was changed
6	Anomaly in RO feed water level meter and UF feed water level meter	Value of LIT301 & LIT401 was changed

6. Summary of Results

7.1 iTrust ADMs

Table 4 shows the anomalies detected by each iTrust detector. A cell marked "X" indicates that the anomaly was detected. Tables 5 and 6 detail the performances of iTrust ADMs (anomalies detected, number of alarms generated, and the number of true positives (TP), false positives (FP) and false negatives (FN)), in numbers and percentages, respectively. The

rate of TP and FP are in relation to total alarms generated and the rate of FN is in relation to the total number of anomalies.

Table 4: Detection of anomalies by iTrust ADMs

Anomaly S/N*	DAD	AEGIS	AICrit	HybMon	Attester
Aug3AM					
1	X	X	X		X
2	X		X		X
3	X	X			
4		X	X		X
5					
6			X		X
7	X	X	X		x
8	X				X
9			X		X
10					X
11	X	X	X		
12	X	X			X
13	X	X	X		X
14	X	X	X		X
15		X			X
16	X				
17	X				X
18	X		X		X
19					
20	X	X	X		X
21					X
22					X
23	X	X			X
24	X	X	X		X
25	X	X	X		
26	X	X	X		
Aug6AM					
1	X	X	X		
2		X	X		
3	X	X	X	X	X
4		X	X	X	X
5	X	X	X	X	
6	X	X	X	X	

*Anomaly S/N: please refer to Table 2

Table 5: Performance of iTrust ADMs, number

Detector	Session									
	Aug 3AM					Aug 6AM				
	Total # Anomalies	Total # Alarms	TP	FP	FN	Total # Anomalies	Total # Alarms	TP	FP	FN
AEGIS	26	7450	958	6483	12	6	998	137	861	0
AICrit	26	15812	1998	13814	12	6	1990	38	1952	0
DAD	26	761	761	0	9	6	130	130	0	2
HybMon	26	2513	0	2513	25	6	563	23	540	2
Attester	26	168	83	85	7	6	1821	131	1690	4

Table 6: Performance of iTrust ADM, %

Detector	Session					
	Aug 3AM			Aug 6AM		
	TP	FP	FN	TP	FP	FN
AEGIS	12.86%	87.02%	46.15%	13.73%	86.27%	0.00 %
AICrit	12.64%	87.36%	46.15%	1.91%	98.09%	0.00 %
DAD	100.00%	0.00%	34.62%	100.00%	0.00%	33.33%
HybMon	0.00%	100.00%	96.15%	4.09%	95.91%	33.33%
Attester	49.40%	50.60%	26.92%	7.19%	92.81%	66.67%

7.2 EBTs ADMs

Table 7 shows the anomalies detected by each EBTs' detector. A cell marked "X" indicates that the anomaly was detected. Tables 8 and 9 detail the performances of EBTs' ADMs (anomalies detected, number of alarms generated, and the number of true positives (TP), false positives (FP) and false negatives (FN)), in numbers and percentages, respectively. The rate of TP and FP are in relation to total alarms generated and the rate of FN is in relation to the total number of anomalies.

Table 7: Detection of anomalies by EBTs

Anomaly S/N*	EBT01	EBT02	EBT03	EBT04	EBT05
AUG3AM					
1		X	X		
2		X			
3				X	X
4		X			
5		X	X		
6		X	X	X	
7		X	X	X	
8		X			
9		X		X	
10		X		X	
11		X			
12		X		X	
13		X		X	
14		X		X	
15		X			
16					
17		X			
18		X		X	
19					
20		X		X	
21		X			
22		X	X		
23		X	X		X
24			X		
25		X	X		
26			X		X
Aug3AM					
1		X	X		X
2		X	X	X	X
3		X	X		
4		X		X	
5		X			
6					

*Anomaly S/N: please refer to Table 2

Table 8: Performance of EBTs ADMs, number

Detector	Session											
	Aug 3AM						Aug 6AM					
	Total Anomaly	Total Alarms	Total OT Alarms	TP	FP	FN	Total Anomaly	Total Alarms	Total OT Alarms	TP	FP	FN
EBT01	26	0	0	0	0	26	6	0	0	0	0	6
EBT02	26	1506	529	193	336	5	6	616	16	16	0	1
EBT03	26	3773	3655	720	2935	16	6	57	53	9	44	2
EBT04	26	98	96	76	18	16	6	16	10	3	7	3
EBT05	26	17	17	10	7	23	6	2	2	2	0	4

Table 9: Performance of EBTs ADMs, %

Detector	Session					
	Aug 3AM			Aug 6AM		
	TP	FP	FN	TP	FP	FN
EBT01	0.00%	0.00%	100.00%	0.00%	0.00%	100.00%
EBT02	36.48%	63.52%	19.23%	100.00%	0.00%	16.67%
EBT03	19.70%	80.30%	61.54%	16.98%	83.02%	33.33%
EBT04	79.17%	18.75%	61.54%	30.00%	70.00%	50.00%
EBT05	58.82%	41.18%	88.46%	100.00%	0.00%	66.67%

7.3 Outcomes

8.3.1. Aug3AM session

Across all iTrust ADMs, Attester detected the most number of anomalies (19 out of 26) followed by DAD (17 out of 26.) However, it must be noted that DAD has a FP rate of 0% compared to Attester's FP rate (50.6%.) Hybmon failed to detect any anomalies resulting in all alarms raised as FP. AICrit and AEGIS raised huge number of alarms (7,450 and 15,812 respectively); however, only about 12.5% of the alarms for (each of the detector) turned out to be true positives. Out of the total 26 anomalies discovered, AICrit and AEGIS detected 14 anomalies during this session.

Across all EBTs ADMs, EBT02 detected the most number of anomalies at 21 out of 26, followed by EBT03 and EBT04 (both with 10 out of 26.) Between EBT03 and EBT04, EBT04 fared better as it had a higher TP rate and a lower FP rate.

8.3.2. Aug6AM session

Across all iTrust ADMs, DAD performed the best as it had the highest percentage of TP at 100.00% with no FP. However, DAD failed to detect 2 of the 6 anomalies. AICrit and AEGIS detected all 6 anomalies but reported a significant number of false positive. HybMon managed to detect 4 out of 6 anomalies and Attester detected 2 out of 6 anomalies.

Across all EBTs, both EBT02 and EBT05 had a TP rate of 100%. However, EBT05 has a higher FN rate compared to EBT02. EBT02 has the most anomalies detected (5 out of 6), followed by EBT03 (4 out of 6) and EBT04 (3 out of 6.)

7.4 Improved iTrust ADMs

iTrust blue teams were given the opportunity to update their ADMs to improve their algorithm after the event. The improved ADMs can then be used to re-evaluated against the event's dataset. The evaluations were done by the developer. Below shows the updated evaluation for AICrit.

Table 10: Updated evaluation for improved AICrit

Detector	Session									
	Aug 3AM					Aug 6AM				
	Total # Anomalies	Total # Alarms	TP	FP	FN	Total # Anomalies	Total # Alarms	TP	FP	FN
AICrit	26	2118	1998	108	12	6	52	38	14	0

9. Conclusion

This year, CISS was held online and invited more participants than in the past from various countries from different backgrounds. A total of thirteen Red Teams and six Blue Teams (five commercial products and an iTrust Blue Team featuring five ADMs) took part in this event. The Red Teams had a unique opportunity to attack a realistic process plant to cause process anomalies. The attacks allowed us to understand composite Tactics, Techniques and Procedures (TTPs) that can be used for enhanced Operation Security (OpSec).

The attacks were scored based on target selection, mode of entry, attack target, attack success and detection. The Blue Teams showcased their ability to detect such sophisticated attacks and by using CISS as a platform to validate and assess the effectiveness of their technologies. Blue teams sent live alarms to the alert logging system. These logs were then analysed by internal iTrust team after the event and scored based on detection of process anomalies in terms of TP, FP and FN. Importantly, CISS has developed capabilities for defending critical infrastructure under emergency such as cyber-attacks which is crucial for safeguarding our interests.

10. Acknowledgements

iTrust sincerely thanks the numerous members of red and blue teams who found time to participate in the first ever international on-line cyber-security exercise focused on exclusively on physical critical infrastructure. Thanks also to the judges who found time to judge this gruelling event.

11. Nomenclature

Alert Logger: Automates the process of logging and sending time-stamped alerts by blue teams to iTrust

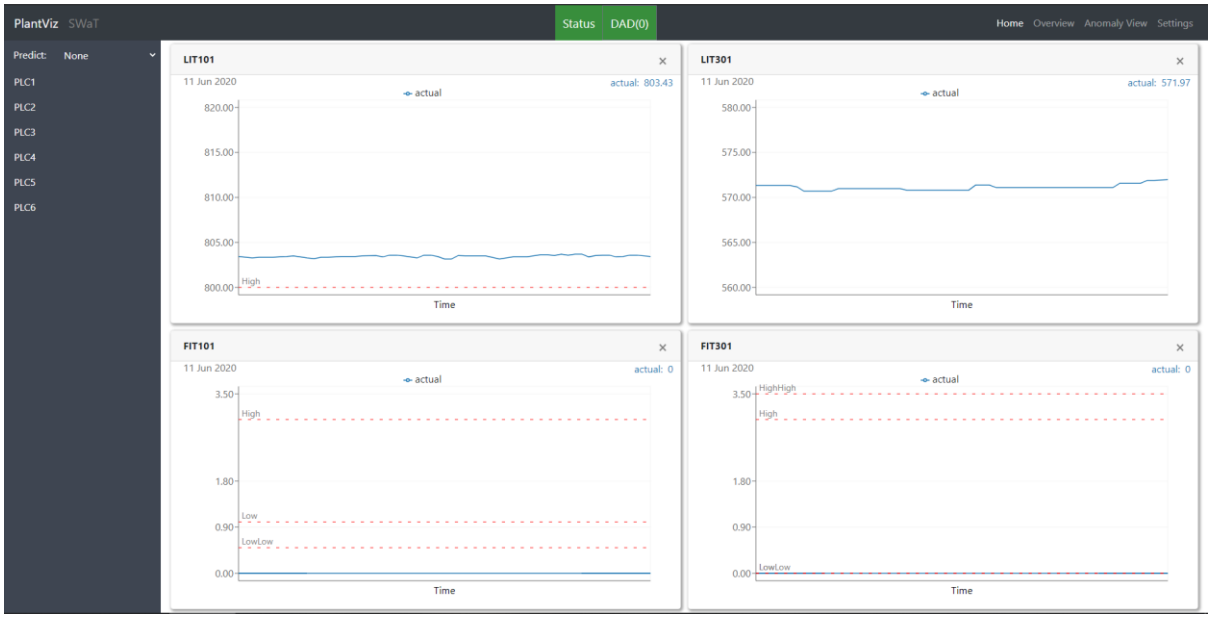
Attack Launcher: Optional platform for red teams to select and launch attacks

Attack Logger: Communicate attack intentions and steps to White Team and judges, and log attacks as they occur

PEPPR: Collection of tools (PlantPlayer (PP), PlantViz (PV) and PlantIO (PI)) that allows playback of historical data to enable blue teams to test their own detection systems

PEPPR-PP: Tool to playback past historical data

PEPPR-PV: Visualisation tool for live, prediction, and anomaly data from detectors.



PEPPR-PI: A suite of tools that allows detectors to save, publish and playback live, the predicted plant state and alerts. The data which is saved or published can be used by other tools on the same network.

<End of document>