CISS 2023 Report

Organised by:

The Digital and Intelligence Service, Ministry of Defence, Singapore
iTrust, Centre of Research in Cyber Security, SUTD, Singapore

Sponsored by:

Cyber Security Agency of Singapore

Date:

17 Aug to 24 Aug 2023

Report by:

Mr Francisco FURTADO, Mr Jonathan Tay, Mr TEO Xu Kai, Mr TEO You Xiang, Mr Tiger LIM, Mr Ivan CHRISTIAN, ME4 Benedict TAN

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Executive Summary

The finals of the seventh run of iTrust's international technology assessment exercise, the Critical Infrastructure Security Showdown 2023 (CISS 2023), was held from 17 to 24 August 2023 at the Singapore University of Technology and Design (SUTD) with the objectives of improving the understanding of composite Tactics, Techniques, and Procedures (TTP) for enhanced operation security, validating and assessing the effectiveness of technologies developed by researchers associated with iTrust[1], and developing capabilities for defending critical infrastructure against cyber-attacks.

The competition was held in two stages, with Stage 1 being a Capture the Flag (CTF) event where 41 Red Teams competed. The top 10 Red Teams from Stage 1 advanced to the final stage, where they were given specific attack objectives to achieve and points were awarded. This year, the judges activated a Wild Card to add an 11th team to the Finals. The Wild Card was used for Team OPENEYES as they were the first team to complete all the OT challenges.

The Intrusion Detection System (IDS) Teams, composed of iTrust's anomaly detectors, and three commercial products, were installed to detect the anomalies resulting from the attacks. The IDS Teams were only present in the final stage of the competition and were tasked with detecting the anomalies. For the evaluation of the IDS Teams, a new 5-metric evaluation framework was introduced: correctness, explainability, accuracy, responsivity and disruptivity. This framework is still in its infancy and its main objective is to provide a holistic and quantitative approach to evaluating the effectiveness of an IDS in assisting incident response.

Overall, the evaluation of both Red Teams and IDS Teams highlights the importance of continuous research and development in the field of ICS security. These evaluations provide valuable insights for improving the performance of both offensive and defensive cybersecurity measures, which are crucial for the protection of critical infrastructure.

In conclusion, CISS 2023 was a successful event made possible with the support of DIS and CSA. We are grateful for their contributions and look forward to continued partnerships in future initiatives.

2. Exercise Overview

CISS 2023 was the seventh run of iTrust's international technology assessment exercise. The objectives were to enhance the understanding of composite TTPs for operation security, validate and assess the effectiveness of technologies developed by researchers associated with iTrust, and develop capabilities for defending critical infrastructure against cyber-attacks.

[1] These technologies include automatically generated anomaly detectors using both design and data centric approaches, protection against plant damage, and tools to assist with incidence response.

The competition was held in two stages:

1. Stage 1: CTF event - In this stage, 41 Red Teams, each with up to 8 members, competed against each other in a 48-hour CTF competition. The top 10 teams from Stage 1 advanced to the final stage. The judges activated a Wild Card to add an 11th team to the Finals. The Wild Card was used for Team OPENEYES as they were the first team to complete all the OT challenges.

2. Final Stage - In the final stage, the top 11 Red Teams from Stage 1 were given specific attack objectives to achieve. Points were awarded based on the success of the Red Teams in achieving their objectives. The top three Red Teams received cash prizes of S$4,000, S$2,000 and S$1,000 respectively. The IDS Teams were composed of three commercial intrusion detection products and two iTrust anomaly detectors. They were tasked with detecting the attacks launched by the iTrust Red Team.

The event was co-organised by the Ministry of Defence (MINDEF) Singapore and sponsored by the Cyber Security Agency (CSA) of Singapore. The results of CISS 2023 provided valuable insights into the current state of Operational Technology (OT) cybersecurity technology and capabilities and highlighted areas for improvement in TTP and technology development.

Overall, the Critical Infrastructure Security Showdown 2023 (CISS 2023) was a successful event that helped to further the understanding of composite TTP for enhanced operation security, validated the effectiveness of technologies developed by researchers associated with iTrust, and developed the capabilities for defending critical infrastructure against cyber-attacks.

3. Methodology

The CISS 2023 was conducted in two stages: Stage 1 and the Final Stage.

3.1. Stage 1: CTF Event, 11 JUL 0800 hrs to 13 JUL 0759 hrs

Stage 1 was conducted in a Lord of the Rings-themed Jeopardy CTF format using the CTF Platform by iTrust's sister lab, the National Cybersecurity R&D Laboratories (NCL) at the National University of Singapore. Figure 1 shows the CTF logo. The story and scenarios created in Stage 1 following Chapters 1 to 12 of The Fellowship of the Ring. Stage 1 took place from Tuesday, 11 JUL 0800 to Thursday, 13 JUL 0759 hrs (UTC+8). A total of 41 teams, comprised of up to 8 members each, participated, with concurrent 48 hours playtime. The CTF platform was provided by NCL (https://ciss-lotr.ctfd.io/) and admin support and team communication were supported using Discord. The teams were ranked by the number of points they accumulated, with ties being broken by the time of completion. The top 10 teams proceeded to the final round of CISS 2023, which was scheduled for 12 - 24 AUG 2022.

Figure 1: Stage 1 CTF Logo

The challenges of the CTF consisted of various categories, including General Knowledge in Multiple Choice format, Forensics and Industrial Control Systems (ICS) Protocols which required teams to decode or encode, as well as interact with ICS services to read or write. There were also unlockable challenges with undisclosed point values. The General Knowledge category consisted of classic multiple-choice questions, with a maximum of two attempts allowed and the second attempt valued at 25 points. The other categories involved standard questions, with hints curated to be useful. The total possible points per team were 3200 distributed over 30 questions. Figure 2, Figure 3 and Figure 4 show examples of the CTF challenges.

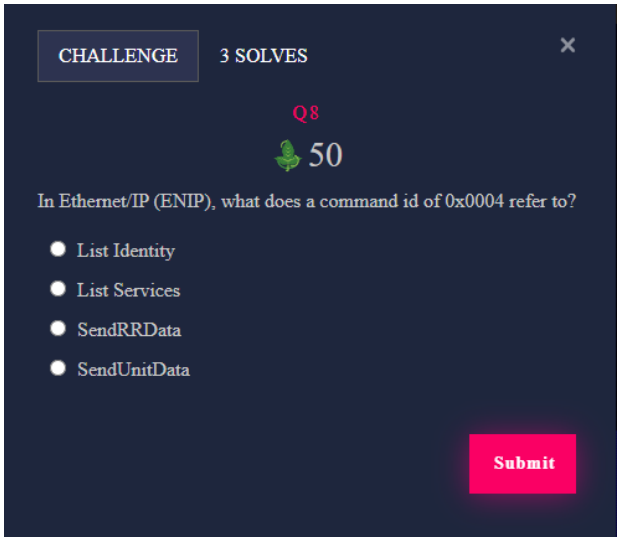

Figure 2: Example of a General Knowledge question.

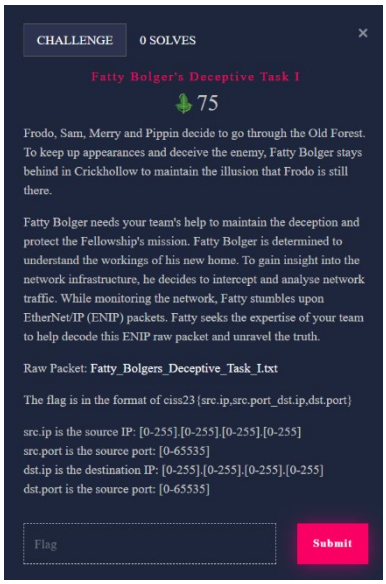

Figure 3: Example of an ICS Protocol Decoding question.

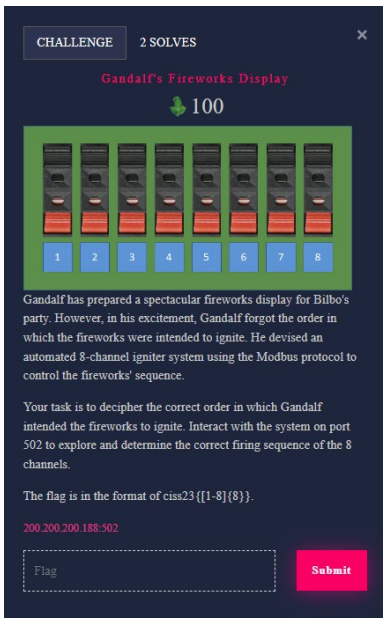

Figure 4: Example of an ICS Protocol communication question.

After 48 hours of gameplay, most of the challenges were solved. Due to the different challenge levels, the CTF event was able to distil teams that were more proficient in Operational Technology. As such, the top 10 teams came with close scores, as shown in Table 1. Being an OT competition, the Judges agreed to activate the Wild Card to add an 11th position to the Finals. The Wild Card was used on OPENEYES as they were the first to complete all the OT challenges. Table 2 shows the top 3 teams that completed all the OT challenges and the time taken.

Table 1: Score and placing of the top 10 teams of the CTF event.

Placing	Team Name	Score
1	ADFCSA2	3900
2	Axe	3900
3	Undecided	3900
4	RED ALERT	3900
5	Shellcode for Cereal	3900
6	404 APT Not Found	3900
7	Bristol Cyber Security Group	3900
8	Yolosw4g	3875
9	COMCYBER FRA	3860
10	UncleCY	3850

Table 2: Time taken to complete OT challenges.

Placing	Team Name	Time to complete OT
1	OPENEYES	11 July, 11:57:30 PM
2	Bristol Cyber Security Group	12 July, 4:13:34 AM
3	404 APT Not Found	12 July, 7:06:32 AM

3.2. Final Stage, 17 AUG to 24 AUG

The final stage took place from Monday, 17 AUG to Friday, 24 AUG over 11 slots. The duration of each slot was 4 hours and was scheduled from 0900 hrs to 1300 hrs or from 1400 hrs to 1800 hrs (GMT+8) daily, with a one-hour break in between for system reset. The top 11 Red Teams were given a set of attack objectives to achieve, while the Blue Teams were tasked with detecting and responding to the simulated attacks. The Blue Teams were composed of iTrust's anomaly detectors and 3 commercial products. Points were awarded to the Red Teams for each objective achieved, while the Blue Teams were evaluated on their ability to detect and respond to the simulated attacks.

3.2.1. Exercise Scenario

The Finals scenario continued the theme of the Fellowship of the Ring from Stage 1, based on the journey of the Fellowship through the mines of Moria. Lord Elrond and his Rivendell scouts sent over intel reports regarding the mines. The scouts report that much of the mines have been taken over by Sauron's forces. It appears that Sauron has restarted mining operations. Orcs have been transporting Mythril out of the mines, headed straight for Mordor. Based on the information extracted from the orcs, the quickest and safest route from the west gate to the east gate is the following:

West gate - Chamber of Marzabul - Magic Chamber - Water Chamber - East gate

The Fellowship must take this route, or they risk death by Sauron.

The Magic Chamber

The Magic Chamber has two distinct installations that are connected. The first installation is a pair of titanic metal pipes extended from the walls and into two monolithic structures. Sauron is transporting miasma from Gas Wells deep in Caradhras to Moria using these metal pipes.

The second installation is an amalgamation of the two monolithic structures and massive metal spikes, with numerous metal ropes that stretch through the chamber walls. Crimson lightning arcs from one metal rope to the next, faster than the eye can see.

Lord Elrond theorised that the miasma from the metal pipes serves as fuel for the two monolithic structures, which then generates magic. Magic is then channelled into the metal ropes which further powers the Water Chamber, Mining Equipment (Critical Load) and Ore Conveyance (Non-Critical Load).

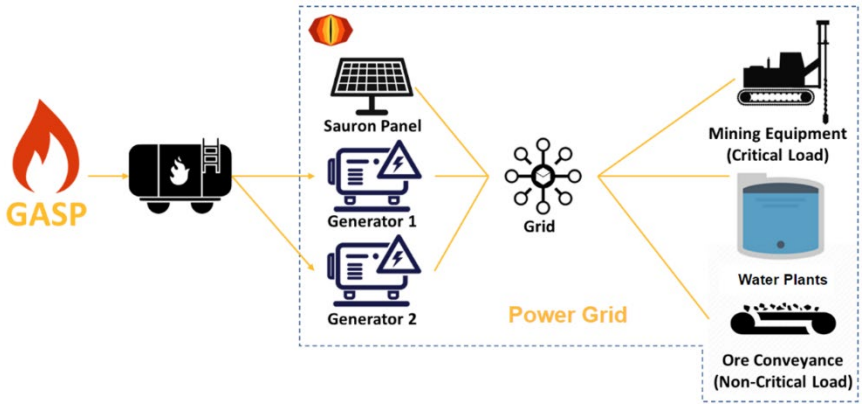

Figure 5: Moria's GASP and Power Grid configuration after Sauron's control.

The Water Chamber

The Water chamber also has two distinct installations that are connected.

The first installation is a vast lake of groundwater siphoned away into a series of interconnected water tanks. Strong-smelling herbs and minerals are added to the water tanks. The water is then channelled into three purification chambers to produce clean water. The scouts report, however, that Orcs cannot drink purified water, as it is poisonous to them. And so, the Orcs have been consuming the rejected purified water instead.

The second installation is a series of four towering water tanks that reach the chamber ceiling. The second installation siphons water from the first installation, taking in clean water and storing it in its tanks. These tanks further supply water to six smaller water tanks within the mines. The six tanks all carry different functions, from the cooling of mining equipment to the hydraulic operation of gates and bridges.

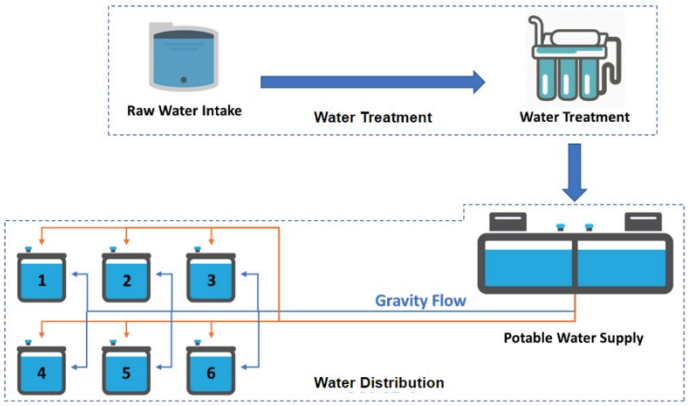

Figure 6: Moria's Water treatment and distribution plants after Sauron's control.

The orcs configured the tanks as follows. Tanks 1 and 2 supply water for cooling of mine equipment, tank 3 for sanitation, tank 4 for ore extraction/conveyance and finally, tanks 5 and 6 for gates and bridges hydraulic operation.

3.2.2. Red Team Activities

As part of the intelligence and pre-attack phase, red teams were provided with manuals, packet capture and process data of all the systems one week before their timeslot. These files provided hints into the protocols, IP addresses and ports that were active in the network.

Before the start of their timeslot, each red team was given 20 minutes to join the Zoom Room and connect to the FUA using the Virtual Private Network (VPN) credentials provided. During this time, teams are also allowed to install any additional tools they deem necessary for the competition.

The communication and coordination between the Red Team and the judges were conducted through the Red Team Lead, and all team members were required to share their screens. The recording of the screen was for analysis purposes and will not be published or shared without the team's permission.

Each team was given the attack objectives one week before their timeslot and given a total of 4 hours which began when the Judge declared "Begin." There was no limit on the number of sessions or explicit permission from the judges for enumeration. To score, the team must declare the objective to achieve and explain how they plan to launch the attack. This required explicit permission from the judges.

3.2.3. Scoring Matrix for Red Teams

The Red Teams were assessed in real-time by a team of judges comprising cybersecurity experts and engineers working in the critical infrastructure domain. The judges scored each team based on the attack objectives achieved by the teams.

The total score, S, for each team, was computed based on two factors:

$$Total\ score, S_i = A_i + (A_i * B_i)$$

Where: -

- A is the sum of attack objectives scored.
- B is the bonus score (%) based on rank R using Table 3.

For the bonus score, the teams were ranked against their peers where the total signatures detected and number of packets generated by each team were taken into account. The ranks were scored using the equation below.

$$R_i = \left(\frac{S_i}{\sum_{n=1}^{11} S_n} * 0.7 + \frac{F_i}{\sum_{n=1}^{11} F_n} * 0.3\right)$$

Where: -

- S = signatures detected
- F = number of packets generated

Table 3: Rank to bonus table.

Rank	1	2	3	4	5	6	7	8	9	10	11
Bonus	11%	10%	9%	8%	7%	6%	5%	4%	3%	2%	1%

3.2.4. IDS Teams Activities

The IDS Teams were tasked with installing their systems into the network to ingest various inputs like network packets or historian data from the exercise platform. The network was provided through an Ethernet cable, with an average throughput of 700 Mbps while the Historian data was a live feed of the latest row from the database. The systems generated syslogs which were sent to a Graylog collector for analysis. Between 26 JUL and 16 AUG, the IDS teams were provided with 4 sessions of 4 hours each to perform baselining and troubleshooting activities to ensure that their systems were functioning correctly. IDS Teams were also provided with 1 session of 2 hours of active network scanning.

During the Finals, the IDS Teams were given the option to monitor and analyse their logs. A special Red Team was assembled to launch specific attacks on 28 AUG, and the logs generated by the IDS teams were subsequently analysed.

In total, there were four IDS teams including 3 from commercial vendors and 1 from Institutes of Higher Learning (IHLs). Only two of the commercial teams were finally evaluated.

4. Red Team Analysis

4.1. Tools Used

4.1.1. Enumeration

The red teams made use of the usual tools for Enumeration such as Nmap though new tools such as Fscan, RustScan and Netdiscover were seen being utilized by red teams. Allowing them to map the networks with more detail and speed.

Table 4: Tools used for Enumeration.

Tool	Description
Nmap	The standard tool used for Enumeration. It can do simple operations such as ping sweeps and port scanning to more complex operations such as vulnerability scans. Nmap Scripting Engine (NSE) can be utilized to further expand the tool's capabilities
Fscan	Made by shadow1ng, Fscan is capable of conducting comprehensive scans on the host machines. Allowing red teams to gather important information such as details of the Network Interface Card (NIC) of the machine and even credentials of various services it might be running quickly.
RustScan	RustScan is capable of scanning all 65535 ports of a machine in seconds. Red teams will be able to get all ports that are open on a machine much faster than other tools. It also has a scripting engine to expand its capabilities. Results from RustScan can also be piped into Nmap for further analysis.
NetDiscover	NetDiscover enumerates live hosts in a network by scanning for ARP requests. Passive scans can be done where it only listens for ARP traffic, making it useful in cases where stealth is required.
Feroxbuster	Feroxbuster is a recursive content discovery tool that allows red teams to quickly enumerate sites that are hosted on a web server. A dictionary of known directories can also be used to quickly search for more specific pages.
enum4linux	Enum4linux is used to enumerate Windows and Samba hosts which can give information such as the usernames and OS version.
tshark/ Wireshark	These tools allow red teams to sniff and analyse the network for any interesting traffic. The .pcap files can then be piped into other tools such as GrassMarlin and NetworkMiner where they can be mapped into a visual topology.
FUFF	This helps red teams to enumerate directories, discover virtual hosts or brute-force web applications. It mainly uses the concept of trying many known vulnerable inputs with a web application to determine if any of the inputs compromised the web application.
Vulnscan	Vulnscan is a module that enhances from nmap -sV that uses static binary scanners to detect any vulnerable host .exe files or applications may contain any vulnerabilities.
crackmapexec	Crackmapexec is a post-exploitation tool used to assess large Active Directory networks using SMB, LDAP or WINRM protocol.
GrassMarlin	GrassMarlin provides a method for discovering and cataloguing Supervisory Control & Data Acquisition (SCADA) and Industrial Control System (ICS) hosts on IP-based networks.
NetworkMiner	NetworkMiner Is a network Forensic Analysis Tool (NFAT) be used as a passive network sniffer/packet capturing tool to detect operating systems, sessions, hostnames, open ports etc. without putting any traffic on the network.
Nuclei	Nuclei is a fast network scanner as well as a vulnerability scanner that can be customized to fit different scans on different networks.

4.1.2. Lateral Movement

Most red teams used Proxychains through SSH tunnelling to pivot into the various OT networks. However, as Proxychains is only capable of proxying TCP and UDP packets, it was harder for red teams to enumerate the SWaT/WADI network.

If implemented correctly, teams using Ligolo will be able to pivot into any network with almost zero limitations in terms of protocols. The main drawback would be that an agent is required to be installed on all jump hosts and have its traffic routed properly, adding potential complications.

Table 5: Tools used for Lateral Movement.

Tool	Description
Proxychains	Proxychains can be used to proxy traffic through various types of tunnels such as SSH and Meterpreter sessions. This allows red teams to pivot into networks with little to no configuration or third-party tools. However, the type of protocols that it can proxy is limited which can cause issues with using certain tools further up the cyber kill chain.
Ligolo	Ligolo can forward most types of traffic through a TLS tunnel that is created by its agents. Initial setup can be a little complicated as IP routes have to be created on the red team's machine to properly pipe traffic to the correct interface. As an agent needs to be deployed on all jump hosts, pivoting using Ligolo might be more difficult than creating a simple SSH tunnel and proxy traffic using Proxychains.
sshuttle	Sshuttle acts more like a VPN than a proxy where specific networks will be specified to be forwarded through sshuttle's SSH tunnels. It has similar limitations to proxychains but it is a lot simpler to configure the tunnels.
Chisel	Chisel is a TCP/UDP tunnel that transports over HTTP and is secured via SSH. It is mainly useful for bypassing through firewalls or can also provide a secure endpoint into the network.
Xvncviewer/ tightvncviewer	XVNCViewer is a tool to connect to any compatible VNC systems, to take control of the desktop environment.
xrdp	XRDP is a tool used to connect to RDP machines, providing a fully functional remote desktop.
remmina	Remmina is a remote desktop client that supports POSIX-based systems as well, supporting RDP, VNC and more protocols.
ngrok	Ngrok is a web hosting tool where attackers can temporarily create a web server to download post-exploitation tools into the victim's machine
xfreerdp	Xfreerdp is a remote desktop tool used to connect to RDP machines.

4.1.3. Exploitation of IT Network

BlueKeep (CVE- 2019-0708) is found on one of the jump hosts and if exploited successfully, gives them access to the SWaT/WADI network. Metasploit has a module to exploit that vulnerability, but a correct GROOMBASE value must be provided else the jump host will crash upon exploitation. The module provides a few templates that include the GROOMBASE value depending on the environment that the Operating System (OS) is running on. However, there are no templates for the specific environment for that jump host. Thus, the value must be provided by the red team which might not be possible to find within the given timeframe.

The second entry point into the SWaT/WADI network would be through a misconfigured MS-SQL server running Windows Server 2012. They will first need to gain the credentials by

viewing CCTV footage that can be found on the EPIC network. The credentials can then be used to log into the MS-SQL service and execute shell commands using xp_cmdshell where they can use it to download and execute payloads on the MS-SQL server.

An alternative way to enter the second entry point would be to make use of Windows Remote Management (WinRM). After providing the credentials that they have found, they can make use of tools such as evil-winrm to execute commands. This route can be easier for the red teams as they will only need to brute force the Administrator password, a user that exists on all machines running Windows Server.

Table 6: Tools used for Exploitation of IT network.

Tool	Description
Impacket	Impacket is a suite of tools written in Python that provides access to packet and protocol manipulations. There is an extensive collection of pre-made scripts that can help red teams to connect to various services. For CISS 2023, *mssqlclient, psexec, smbclient, rpcdump and samrdump* are used by the red teams.
HavocFramework	Havoc is a post-exploitation tool C2 that is meant to control multiple victim machines at the same time.
Pwncat	Pwncat is a fancy post-exploitation tool for bind shell or reverse shell etc.
Msfconsole/ Metasploit	MSFConsole is an interface to make use of the MSF Framework which includes multiple exploitation/enumeration scripts which allow users to remotely control multiple machines at once.
BurpSuite	BurpSuite is a web vulnerability scanner that allows attackers to modify requests and send malformed requests to the web server.
crowbar	Crowbars is a brute forcing tool that can be used to brute force protocol such as SSH in different manners from the other tools such as using SSH keys instead of SSH username/password.
LinPEAS	LinPEAS is a post-exploitation tool used to enumerate the Linux system. It helps to find vulnerabilities to perform privilege escalation or find other hosts that could be connected to the host.
feroxbuster	FeroxBuster is a web directory brute force tool, that enumerates the different resources that could exist on the web server.
busybox	BusyBox is used to combine commonly used Linux tools into 1 binary file such that it will be easily accessible on victim machines even if they do not have those commands originally.

5. Evaluation of IDS Teams

The IDS Teams were evaluated based on the total OT anomalies detected on the exercise platform using attacks launched by the iTrust Team on 28 Aug. It was selected as it contained the greatest number of OT attacks. A list of attacks is enumerated in Table 7. IT anomalies were not considered as part of the evaluation.

Table 7: Attacks used to analyse the detectors.

Attack ID	Attack Description
A0	Attacker scans and compromises VM
A1	Attackers launch reverse shell setup to gain shell access through Ligolo
A2	Load EPIC Attack Desk Web GUI on the browser
A3	Attacker opens Generator Motor 1 CB (Q2C)

A4	Attacker closes Generator Motor 1 CB (Q2C)
A5	Attackers add the non-critical load by 2.5kW
A6	Attacker reverts the non-critical load values
A7	Attacker pivots to SWaT/WaDi Network through MS SQL Server
A8	Attacker sends pre-made executables (Turn off UV) to MS SQL Server
A9	Attacker executes the executable (Turn off UV)
A10	Attacker stops the executable (Turn off UV) and resets the plant state
A11	Attacker changes LIT101 L: 300 and H: 850 setpoints
A12	Attacker attacks by opening P201, P202, P203 from the attack desk
A13	Attacker stops the attack on P201, P202, P203 and reset the plant state
A14	Attacker reverts changes to LIT101 L (500) and H (800) setpoints
A15	Attacker attacks LIT301 by spoofing the value to 1000 from the attack desk
A16	Attacker stops the attack on LIT301 and resets the plant state
A17	Attacker manually closes 2-MV-005 to stop the water from going to consumers through gravity
A18	Attacker executes attack to close (0%) 2-MCV-201 from the attack desk
A19	Attacker executes attack to spoof 1-FIT-001 value to 200 through the attack desk
A20	Attacker manually opens the switch to cut off power to the water testbed (Q3-2)
A21	Attacker manually closes the switch to restore power to Water Testbed (Q3-2)

5.1. Performance Analysis and Summary

For CISS 2023, a performance metric was introduced to quantitatively measure the performance of the detectors deployed in the exercise. The following subsections describe each metric while Table 8 and Table 9 show how points are awarded to each metric used.

5.1.1. Correctness

Correctness assesses the ability to pinpoint anomalous behaviour. This assessment relies on four crucial identifiers: the stage of the physical process affected; the involved components; the anomaly type; and the underlying cause.

For example, in the context where water unexpectedly drains from Tank 101, a well-performing IDS would be able to pinpoint the issue to Stage 1, in component Tank 101, with the problem being a possible physical leak in the tank and that a valve was opened to drain the water. This detailed understanding, enabled by the four identifiers, allows for swift and targeted corrective action.

5.1.2. Explainability

Explainability assesses the clarity of the alert to convey the nature and severity of threats. These alerts are categorised based on data source: IT/OT (combining network traffic and physical data) and pure OT (based on physical data). In this component, the performance is measured mainly using the Flesch-Kincaid scoring system (FK score), which is a quantitative scoring measure of how easily a sentence, an alert in the context of this category, can be understood. The higher the score, the easier the sentence is understood. An example of this would be “There is a problem with P101” which would result in a score of 99.2.

In the Tank 101 scenario, an IT/OT alert might mention "unusual network activity from Tank 101 control system suggesting unauthorized valve access," further aiding rapid decision-making.

5.1.3. Accuracy

Accuracy assesses the ability to correctly identify real anomalous behaviour coming from the plant. For a solution to be considered stellar, the detector mustn't generate any false alarms as it would be costly and unnecessary to inspect.

For example, a detector observing the behaviour of actuators P101, which is the main pump, and P102, which is the backup pump, in Tank 101 should be considered performing well if it can detect a logical or controller fault when both actuators are open during active plant operations correctly. It should not however raise an alert when P101 closes and P102 opens when it is a valid operation.

5.1.4. Responsiveness

Responsiveness assesses the reaction time in reporting anomalous behaviour to ensure that the incident can be responded to quickly and effectively. As incident response is a critical component in OT security, the shorter reaction time means that there is more chance for the operators to save the plant from permanent damage.

For example, there is an attack on the actuator P101 which forces the pump to continue running even when there is no more water in the previous tank. A well-performing detector should be able to raise an alert as soon as there is a violation in the logic so there is no permanent damage from the pump being dry ran.

5.1.5. Disruptivity

Disruptivity assesses the degree to which the solution disrupts the plant operation. This measurement is done to show the anomaly detector does not affect the normal operation of the plant as it should not be the reason why the plant is behaving abnormally.

For example, a detector should not be sending out network packets that would contribute to the network traffic of the plant and add to the communications bandwidth. The detector should also not create any disturbances in the communications between the plant components such that it would violate the logical sequence of the plant.

Table 8: IDS Performance Metrics Part A

Score	Correctness	Explainability	
		IT/OT	OT
4	Identified - Specific Stage - Specific Component - Component failure type - Reason for failure	Identified - Severity identified - Protocol identified - SIP and DIP identified FK Score: 30.0 – 100.0	FK Score: 70.0 – 100.0
3	Identified - Specific Stage - Specific Component - Component failure type	Identified - Severity identified - Protocol identified - SIP and DIP identified	FK Score: 50.0 – 69.9
2	Identified - Specific Stage - Specific Component	Identified - Severity identified - Protocol identified	FK Score: 30.0 – 49.9
1	Identified - Specific Stage	Identified - Severity identified	FK Score: 10.0 – 29.9
0	None identified	None identified	FK Score: 0.0 – 9.9

Table 9: IDS Performance Metrics Part B

Score	Accuracy	Responsiveness	Disruptivity
4	Model Detects - 90.0% – 100% of attacks - 0.0% – 10% of false alarms	Alarm is generated within 0s - 10s of the anomaly	Solution does not disrupt plant operations
3	Model Detects - 70.0% – 89.9% of attacks - 10.1% – 30.0% of false alarms	Alarm is generated within 11s - 30s of the anomaly	Solution disrupts plant operations 1% - 20% of the time
2	Model Detects - 50.0% – 69.9% of attacks - 30.1% – 50.0% of false alarms	Alarm is generated within 31s - 60s of the anomaly	Solution disrupts plant operations 21% - 50% of the time
1	Model Detects - 30.0% – 49.9% of attacks - 50.1% – 70% of false alarms	Alarm is generated within 61s - 120s of the anomaly	Solution disrupts plant operations 51% - 99% of the time
0	Model Detects - 0.0% - 29.9% of attacks - 70.1% – 100% of false alarms	Alarm is generated as part of the forensics or not generated	Solution disrupts plant operations 100% of the time

Table 10: Detection of attacks by IDS Teams.

Attack ID	IDS1	IDS2	IDS 3
A1	YES	YES	NO
A2	YES	NO	NO
A3	YES	YES	NO
A4	YES	YES	NO
A5	YES	NO	NO
A6	YES	NO	NO
A7	YES	NO	NO
A8	YES	NO	NO
A9	YES	NO	NO
A10	YES	NO	NO
A11	YES	NO	NO
A12	YES	YES	YES
A13	YES	YES	YES
A14	YES	YES	YES
A15	YES	NO	YES
A16	YES	NO	YES
A17	YES	NO	YES
A18	YES	NO	NO
A19	YES	NO	NO
A20	YES	NO	NO
A21	YES	NO	NO

Table 11: Performance of IDS summary.

No	Detector name	Detector Type	Total Score	Correct-ness	Explain-ability	Accuracy	Responsivity	Disruptivity
1	IDS1	IT/OT	17	3	4	4	2	4
2	IDS2	IT/OT	9	2	2	0	1	4
3	IDS3	IT/OT	11.5	2	1.5	0	4	4

Table 12: Sample Alerts generated.

No	Detectors	Alert generated
1	IDS1	9814e9de-32a1-4f85-8c73-23ba0ffef2a5 VI:PROC:NEW-VALUE New OT variable value 2023-08-28 14:53:42 New variable value (300, expected value is 500) for variable 192.168.1.10/0/cip-HMI_LIT101/SAL[0] (cip-HMI_LIT101/SAL[0] at 0) 10 00:0c:29:05:63:d6 00:1d:9c:c7:b0:70 192.168.1.235 192.168.1.10 6.0 ethernetip consumer, web_server consumer, producer, web_server FALSE 192.168.1.235 192.168.1.10 TRUE 37334 44818 ubuntu-2.local OTIDS-SWAT 10.10.10.15 tcp TRUE 831a4b6b-c590-421c-baa5-e9fdd0b38d90 FALSE raised_by, n2os_ids, RTU_ID, 0, base_risk, 6, from_id, 192.168.1.235, host, 192.168.1.10, is_dst_node_learned, true, is_dst_public, false, is_dst_reputation_bad, false, is_src_node_learned, true, is_src_public,

		false, is_src_reputation_bad, false, learn_rules, vi variable 192.168.1.10/0/cip-HMI_LIT101/SAL[0] min_value hex:4072c00000000000, to_id, 192.168.1.10, var_key, 192.168.1.10/0/cip-HMI_LIT101/SAL[0], var_origin, consumer, mitre_attack_for_ics, {"source"=>{"types"=>["Engineering Workstation"]}, "destination"=>{"types"=>["Field Controller/RTU/PLC/IED"]}} 2023-08-28 14:53:42 (ip host 192.168.1.235 and ip host 192.168.1.10 and tcp port 37334 and tcp port 44818) or (vlan and ip host 192.168.1.235 and ip host 192.168.1.10 and tcp port 37334 and tcp port 44818) open 8192.168.1.235192.168.1.1091d6af1218a3aeb6b23 FALSE vmx1 New OT variable value TRUE Undefined Undefined computer IO_module 2023-08-28 14:53:42 1 edfe49b0-30f8-4649-ade4-a86133bf3476 VI:PROC:NEW-VAR New OT variable 2023-08-28 15:01:21 New variable on host 192.168.1.30 with protocol ethernetip (192.168.1.30/0/cip-HMI_LIT301/Sim[0]) 10 00:0c:29:bc:06:83 00:1d:9c:c8:bd:f2 192.168.1.231 192.168.1.30 9.0 ethernetip consumer consumer, producer, web_server FALSE 192.168.1.231 192.168.1.30 TRUE 43134 44818 OTIDS-SWAT 10.10.10.15 tcp TRUE ebbb143e-d741-4a2d-97d6-031042ac6667 FALSE raised_by, n2os_alert, RTU_ID, 0, base_risk, 6, delete_rules, vi variable 192.168.1.30/0/cip-HMI_LIT301/Sim[0] :delete, from_id, 192.168.1.231, host, 192.168.1.30, is_dst_node_learned, true, is_dst_public, false, is_dst_reputation_bad, false, is_src_node_learned, false, is_src_public, false, is_src_reputation_bad, false, learn_rules, vi variable 192.168.1.30/0/cip-HMI_LIT301/Sim[0] is_learned true, mitre_attack_for_ics, {"destination"=>{"types"=>["Field Controller/RTU/PLC/IED"]}}, to_id, 192.168.1.30, var_key, 192.168.1.30/0/cip-HMI_LIT301/Sim[0], var_origin, consumer, alert_data, 2023-08-28 15:01:21 open 8192.168.1.231192.168.1.30a87eaf1218a3af26d27 FALSE vmx1 New OT variable TRUE Undefined VLAN_1 - IO_module 2023-08-28 15:01:21 1
2	IDS2	Computer Dell(192.168.1.203)|IO Module Rockwell Automation(192.168.1.60)|64723|44818|28/8/2023 17:00|Connection start (2023.08.28 17:00:21) with duration is not in periods time of allowed policy pending-acknowledged|
3	IDS3	Abnormal network packet to 192.168.1.10

5.2. IDS 1

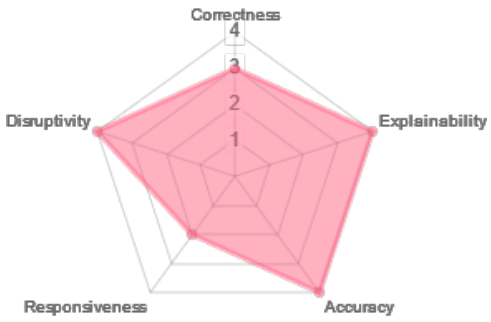

Figure 7: Spider Graph of IDS 1.

Based on the reported evidence, IDS 1 was able to identify which component through the IP address, identify the type of attack, and provide the predicted correct behaviour of the plant.

The identification of the component can be seen in the destination IP address since the IP address would indicate which component is being attacked. The alarm also specifies which incident type it is in the tag "msg". While it is specific to the detector, it describes the type of anomaly that the component is experiencing. The alarm however does not indicate any reason for why the component is failing, be it a component degradation or an anomalous activity.

In the case of explainability, the information disseminated from the alert is clear in terms of the FK score. The score hovers around 80, which indicates that it should be easily understood by educated American 6th-grade students. This means that anyone who has had a 6th-grade education would be able to understand the alerts easily. Additionally, the alerts indicate the source IP addresses and destination IP addresses to show where the anomalies seem to be coming from and which are the target devices. Protocol and severity of the alert are also indicated clearly to show how important and urgent the anomalies are to be resolved. IDS 1 fulfils all the stellar criteria of the performance metric.

Based on the evidence provided in the detection report, IDS 1 was found to detect all the curated attacks. Evidence suggests that the detector was able to detect various types of attacks launched as the curated attacks use various methods to cause anomalies in the plant. As such, with the given evidence, we can see that the detector has achieved the necessary standards of a stellar detector in terms of accuracy.

The detector responds to the anomalies within 60 seconds, as seen by the evidence given in the report. While this standard is different for different types of plants, 1 minute to generate an alert for a water treatment plant requires a quicker response time. As such, we can see that this detector has performed well enough in the standards of the performance metric.

5.3. IDS 2

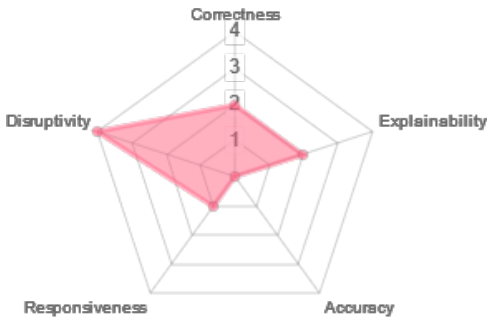

Figure 8: Spider Graph of IDS 2.

For IDS 2, the detector seems to be identifying only whether the alert shows whether the anomaly violates the allowed policy without giving any details on what components and/or policies are violated. The alert does not seem to mention specifics of which part of the control logic is being violated.

IDS 2 alerts clearly show that there are not enough specifics mentioned in the evidence given. Although the IP Address, manufacturer and build of the component show which component is being violated, which can be found in a provided manual, we can see that there are no more details given to show that there is something amiss in the process. IDS 2 alerts only show that there is a policy violated based on their baseline standards. Based on

the performance metric, this does not qualify for a 4-point award in the correctness category.

The alerts, given to us were sufficiently short, succinct, and clear in terms of communicating the necessary information. Information seems to be easily digested and clearly delivered as part of the alert. Its FK score hovers around 60-70. In this category, the alerts should easily be readable for an American-educated 8th grader and should be able to transfer information efficiently regarding the problem in the plant. The detectors were also able to identify the source and destination IP addresses and the time when the alerts were generated but not the protocols involved in the communications.

The accuracy of the detector shows that the solution is not performing well as we were provided only with 6 detections out of the 23 attacks launched. It seems that the detector was only able to detect specific types of attacks, such as unallowed connections to PLCs. The detector, however, does not seem to be able to detect problematic commands onto the PLCs as the only alert messages generated were, "Connection is not in allow policies after system's baseline."

There is, however, other evidence to show false alarms that the detector generates. Out of the 712726 alerts, only 669 alerts were connected to the attacks. The other non-attacking alerts seem to be falsely generated on valid and proper communications. This may be because of an issue with the detector baselining during the baselining period. Among the 669 alerts, only 6 standardised attacks were detected.

In the case of the responsiveness, the detector seems to respond on average 2 minutes after the attack was launched. This would still be considered acceptable in making sure that the alerts are sounded when there is an anomaly in a water treatment facility. However, in a different critical infrastructure, this may cause permanent damage that could have been prevented if the alert was sounded earlier. As such, in the context of a water treatment plant, this detector has performed admirably but not well enough for this category.

In terms of disruptivity, there seem to be no issues related to IDS 1 since the detector is installed on an edge network monitoring the condition of the plant.

5.4. IDS 3

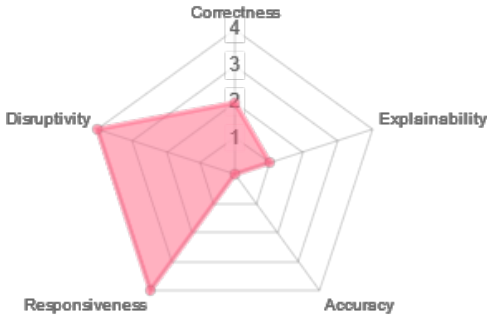

Figure 9: Spider Graph of IDS 3.

The performance of correctness of IDS 3 is like IDS 2 in that it only partially fulfilled the criteria. It identifies the specific component that is failing but does not identify the type of failure and the possible reason for failure.

As for the explainability, the alerts generated short and readable alerts for the operators to read, mainly scoring around 70-80, which is the reading level of an American-educated 7th grader student. There are, however, no details about the severity given to indicate the anomalies' severities. The alerts also only identify the DIP but not the SIP of the malicious actor.

Based on the given evidence, the accuracy of the alerts seems to be less than 50% identified as attacks as only 6/23 attacks were detected. The performance of the detector accuracy seems to have fallen in detecting the attacks. The detector only detected attacks in SWaT as only data from SWaT was fed to IDS 3. This may be because the detector itself is only designed for SWaT specifically. Taking into account the attacks in SWaT, IDS 3 was able to detect 6 out of 9 attacks.

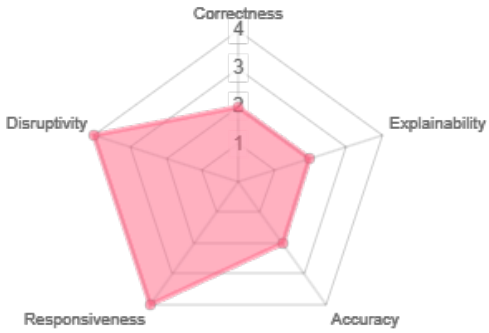

Figure 10: Spider Graph of IDS 3 (SWat Only)

In terms of responsiveness, IDS 3 detector has performed admirably as it was able to generate alerts within 15s of the anomaly happening. Therefore, based on the set criteria, the detector has performed better than the other detectors.

Like the other IDS solutions, IDS 3 was installed as an observer of the plant activity. Therefore, it does not interfere with any of the plant's functionality.

<End of document>