T811: DATA FROM INFORMATION REPOSITORIES

PURPOSE

This Recipe, based upon use of the CyOTE methodology[1] (Figure 1), provides asset owners and operators (AOO) with general guidance for confirming suspicion of the Data from Information Repositories attack technique for the Evasion and Impair Process Control tactics as defined by the MITRE ATT&CK® for Industrial Control Systems (ICS) framework[2,3] allowing them to be able to make informed business decisions based on collaborative analysis of the nature and context of the attack. This document also includes supplemental material with suggestions and recommendations for securing assets and improving detection capabilities. Additional information on this technique can be found in the *Data from Information Repositories (T811) Technique Detection Capability Sheet* for the Evasion and Impair Process Control tactics.[4]

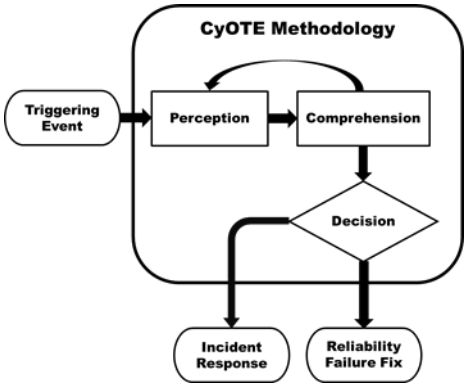

Figure 1: CyOTE Methodology Diagram

POTENTIAL ATTACK TARGETS

Data from information repositories such as historians, file shares, and other process data storage locations provide attackers with information about the environment. This data can include sensitive information such as specifications, schematics, or diagrams of control system layouts, devices, and processes. As defined by the MITRE ATT&CK® for ICS framework, adversaries can use the data collected from this technique for cases of industrial or corporate espionage, or they could leverage the data for longer-term targeting. Examples of target information repositories include reference databases and local machines in the process environment. These data repositories may be located on control servers, data historians, engineering workstations, and human-machine interfaces (HMI).[5]

Advanced attackers might leverage information from the environment to understand the process in order to deliver more damaging effects. Asset owners and operators should consider what information repositories their particular control system contains.

[1] CESER, Methodology for Cybersecurity in Operational Technology Environments, CyOTE Program. Department of Energy, FY21, https://inl.gov/wp-content/uploads/2021/09/CyOTE-Methodology_2021.pdf.
[2] MITRE, Data from Information Repositories, 2021. Available online: https://collaborate.mitre.org/attackics/index.php/Technique/T0811.
[3] © 2021 The MITRE Corporation. This work is reproduced and distributed with the permission of The MITRE Corporation.
[4] CESER, Data from Information Repositories (T811) Technique Detection Capability Sheet, CyOTE Program. Department of Energy, FY21.
[5] MITRE, Data from Information Repositories, 2021. Available online: https://collaborate.mitre.org/attackics/index.php/Technique/T0811.

PERCEPTION: IDENTIFYING ANOMALIES

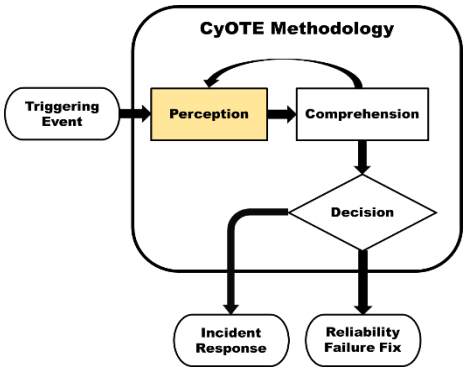

Figure 2: CyOTE Methodology - Perception Step

Perception (Figure 2) is the first active step in employing CyOTE's methodology. CyOTE uses the terms "perception" and "comprehension" as opposed to terms like "detection" and "understanding" for the same reasons that the North American Electric Reliability Corporation (NERC) uses those terms, which were adapted from Dr. Mica Endsley's model of situation awareness[6] – because they cause the reader to think of the necessary individual and organizational human cognition as opposed to merely automated data processing. Perception is defined as a signature capable of being detected by a human was actually detected; perception does not mean opinion or subjective interpretation. It provides the starting point—detection of a triggering event in the organization—for investigation during the comprehension step. This stage involves the identification of observables, anomalies, and triggering events of a potential malicious event; please consult the CyOTE methodology for more information on these terms.[7]

EXAMPLE OBSERVABLES AND ANOMALIES OF THE DATA FROM INFORMATION REPOSITORIES TECHNIQUE

Differences between operational environments require AOOs to understand the generic and unique attributes of the environment under analysis. The anomalies and observables in Table 1 serve as a generic starting point to adapt to the realities of the specific environment. Each potential anomaly includes data sources where it may be observed, as well as what the observables may be.

Note that this table is not intended to be an exhaustive list of anomalies, data sources, or observables tied to the Data from Information Repositories technique.

Table 1: Notional Events

Observables	Anomalies	Data Sources
• Performance degradation during exfiltration from an information repository • Unusual accounts logged in or other signs of unauthorized access	• Significant degradation to a historian or other process datastore • Signs of exfiltration from a historian or other process datastore	Operator or Plant Personnel

[6] Mica R. Endsley, "Situation Awareness Misconceptions and Misunderstandings," Journal of Cognitive Engineering and Decision Making 9, no. 1 (March 2015):4, https://doi.org/10.1177%2F1555343415572631.
[7] CESER, Methodology for Cybersecurity in Operational Technology Environments, CyOTE Program. Department of Energy, FY21, https://inl.gov/wp-content/uploads/2021/09/CyOTE-Methodology_2021.pdf.

Observables	Anomalies	Data Sources
Unusual file storage in information repositories	Unusual file or folder collection on disk with documents	• Operator or Plant Personnel • Windows Event Logs
• Specific Microsoft SysInternals' Sysmon utility event IDs like 5, 7, and 8 • An increase in support desk tickets for the crashed historian and other database-dependent applications • Abnormally high load of queries to an information repository • Failed or anomalous logins for the related service accounts	A process associated with a critical service crashed unexpectedly.	• Windows Event Logs (Enhanced) • Support Desk Tickets
• SELECT queries or other high-risk SQL queries/operations (e.g., DROP TABLE, ALTER TABLE) • Suspicious or anomalous traffic to data repository API ports. This can also be used for MS SQL ports if network volume limits full data collection.	API and SQL queries and writes to/from historians and other operations data stores from unexpected hosts	• Raw Network Data (Captured) • Raw Network Data (Live) • Network Flow Data (Captured) • Network Flow Data (Live) • Application Logs
• SMB file access patterns • Unexpected changes to the SMB share baseline • Changes to or deletion of critical files in SMB share	Unexpected hosts performing unauthorized file requests over server message block (server load balancing [SLB]-based file shares)	• Raw Network Data (Captured) • Raw Network Data (Live)
• Unexpected change in file modification, access, or creation time • Creation of unexpected file metadata or application logs associated with file access	File metadata change (e.g., access time, user) by unauthorized user	• File Metadata
Attempted exfiltration of a decoy or honeypot document	Attempted access, exfiltration, or collection of decoy documents or information	• Application Logs • File Metadata • Raw Network Data (Captured) • Raw Network Data (Live)

STEPS FOR IDENTIFYING HIGH-CONSEQUENCE SYSTEMS, PATHWAYS, AND POTENTIAL TRIGGERS OF DATA FROM INFORMATION REPOSITORIES

Asset owners and operators aiming to develop and implement the Data from Information Repositories technique capability should consider a phased approach to development to include continuous testing

and evaluation throughout its life cycle. To complement this, it is highly encouraged to use the following steps to map out existing OT infrastructure both logically and physically. This supports capability development and the analysis of potential alerts, enabling the quick identification of control devices communicating within the infrastructure.

Continual testing and evaluation will ensure the newly introduced software does not negatively impact, adversely affect, or introduce vulnerabilities into the existing OT environment. As a guideline, during the development phase, secure coding practices should be employed.[8]

1. Identify devices containing configuration files and device history
2. Identify data historians containing event history and other logs
3. Identify what devices and protocols contain sensitive data and should be monitored as targets for this technique
    a. E.g., workstations containing:
        i. Databases
        ii. Device configurations
        iii. Operator software
        iv. Sensitive data
    b. Devices with local databases
    c. Identify parsers for the applicable protocols of each potential trigger
4. Identify where the capability will be located and when it will operate
    a. Example capability locations: from firewall, integrated host, server, IDS/IPS, other
    b. Example operating timeframes: at startup, real-time, daily, weekly
5. Identify tap points (sensors) for observing traffic for identified devices
    a. This includes servers, HMIs, operator workstations, data historians, security appliances, and logging locations (hosts)
        i. Plan sensor placement based on locations within the architecture that provide context related to the anomaly and prioritize systems at greater risk
    b. Monitoring and traffic aggregation may necessitate tap placement on both sides of the identified devices
        i. E.g., MAC addresses may change as information traverses networking infrastructure such as protocol converters
    c. Recommend establishing capture requirements for monitoring OT traffic and their locations[9, 10]
        i. Storage (how much and for how long)
        ii. Line rate (e.g., 1/10/40/100 Gb)
        iii. Live stream data or full Packet Capture (PCAP) offline
        iv. Central versus distributed collection/analysis/alerting

[8] Microsoft, "Security engineering SDL practices," Blog, available online at https://www.microsoft.com/en-us/securityengineering/sdl/practices.
[9] CESER, Security Monitoring Best Practices, CyOTE, 2021.
[10] CESER, Lessons Learned, CyOTE Program, Department of Energy, 2021.

COMPREHENSION: ANALYZING TRIGGERS AND THEIR CONTEXT

The second step of employing the CyOTE methodology is Comprehension (Figure 3). Comprehension is the ability to understand an anomaly in all its relevant context across the operations, OT, information technology (IT), business, and cybersecurity domains. Comprehension involves understanding the nature and possible origins of the anomaly and developing broader awareness of the overall context in which the triggering event came to be. Because of the multidisciplinary approach used for a sufficient investigation, comprehension for the purposes of CyOTE is an organizational ability, not an individual one.

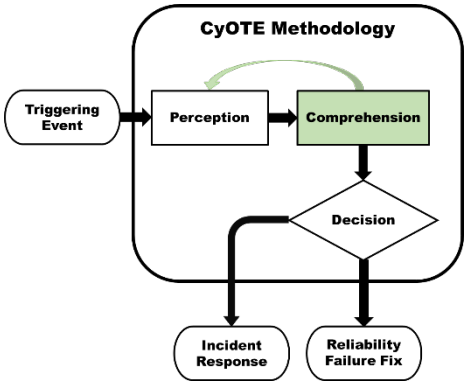

Figure 3: CyOTE Methodology - Comprehension Step

IDENTIFY AND COLLECT AVAILABLE INFORMATION RELATING TO DATA FROM INFORMATION REPOSITORIES

Developing comprehension around an anomaly is most effective and efficient when small core teams of full-time system operators, OT technicians, and cybersecurity analysts from different departments and external resources come together to purposefully focus on the problem in the context of their shared organization. Table 2 includes a list of different organizations that could be leveraged to collect information on an anomaly. Depending on the size and resources of your organization, some of these resources might not be applicable.

Table 2: Business Organizations that Support Information Collection for Data from Information Repositories

Organization	Capacity
• System Operations Departments • Engineering Departments	Control center field operators and real-time engineers should be one of the first sources consulted. Information collected might include manual logs, notes from field personnel investigations, or other records if the anomaly involved a disruption in the system above some established threshold.
Cybersecurity Departments	Includes those responsible for the confidentiality, integrity, and availability of the organization's digital assets provide the threat-informed perspective and bring experience and capabilities to analyze situations and data for security issues.
Original Equipment Manufacturers (OEM)	Includes those who produce and support the hardware and software present within the industrial environment. OEMs might or might not be under support contracts but might

Organization	Capacity
	provide technical documentation and expert advice on expected device behavior.
Third-Party Support Vendors	Cybersecurity, threat intel, and other vendors that provide subject matter expertise might be able to provide insight into anomalies surrounding the trigger conditions.

Since access to raw data typically requires coordination with human organizational oversight, it generally is better to pursue information and context from different departments within the organization. When needed, have them provide the identified data under their control for shared analysis. Appendix A provides specific datasets and information that could be helpful depending on the situation.

STEPS FOR IDENTIFYING AND EXTRACTING INFORMATION FROM OT NETWORK TRAFFIC FOR ANALYSIS OF DATA FROM INFORMATION REPOSITORIES

The information on high-consequence systems, pathways, and potential triggers collected previously becomes more refined for collection in this process. Extracted information should be prioritized based on importance, with timelines established for capturing and holding information for analysis and review.

1. Identify devices and protocols specific to the OT environment to monitor relevant messages
    a. This includes HMIs, operator workstations, data historians, servers, security appliances, and logging locations
2. Establish capture requirements for monitoring OT traffic and their locations
3. Identify applicable triggers
4. Prioritize extracted information based on importance
    a. Establish timelines for capturing and holding information for analysis and review

STEPS FOR ANALYZING EXTRACTED FIELDS AND IDENTIFYING ANOMALIES WITHIN MESSAGES THAT INDICATE DATA ACCESS FROM INFORMATION REPOSITORIES

The suggested fields above are applied to data analysis and assist in establishing triggers. The list below includes suggested ways to use the extracted fields to help identify anomalies that could be alerted on. Any message revealing two or more parameters are given greater precedence for analysis and correlation with other potential triggers to identify potentially anomalous/malicious messages.

1. Identify traffic of interest to alert on
    a. Document command-initiated triggers based on host
        i. Identify the existing traffic origination points

ii. Include the frequency and type of command(s)
   b. Identify and match high-risk command message type(s)
   c. Determine if the alert is valid or invalid based on analysis of the message parameters and source
2. Identify messages trying to access sensitive data:
   a. File access commands
   b. SQL database queries
   c. File read, write, or transfer methods
   d. Found in FTP, DNP3, SEL
   e. Event History access commands
   f. Found in Telnet, SSH, SEL, ICCP
3. Remote access and login events to operator workstations, HMIs, data historians
   a. Found in RDP, SSH, Telnet
4. Identify traffic coming from new or abnormal hosts
   a. Analyze host lists for repository messages issued to end device(s)
   b. Conduct a comparative analysis to identify new connections and alerts versus old ones
5. Establish triggers
   a. Incorporate the analysis findings provided in Step 3 and implement to refine alert parameters to focus on the useful information and minimize the number of non-useful alerts

REPORTING TRIGGERING EVENTS

Security needs to consider the operational and business aspects of the organization. Timely reporting of anomalous activity reduces the time needed to comprehend triggering events and make a decision, leading to reduced process outage times. The following guidance aims to enhance your organization's existing reporting procedures, although these guidelines might also vary based on the composition of your organization. If your industry or organization requires stricter timeframes or processes, you should default to the more stringent timeframe and procedures.

Programmatic alarms should already be logging information, but organizations should document any contextual information for further analysis and reporting. Organizations should also report unusual operating conditions not based on programmatic alarms. Examples of essential details to record include the usual and intended conditions and how the observed operating condition deviates from expected. Table 3 below provides reporting guidelines for a sample of triggering events but is by no means an exhaustive list.

Table 3: Triggering Event Reporting Suggestions

What To Report	Whom To Report To	Recommended Timeframe	Desired Outcome
Significant unexpected increase in information repository queries over the network	• Network security team • Team responsible for network resource	• 1 hour (operational impact) • 48 business hours	• Identify the source(s) associated with the increase in network queries • Identify if a change to the configuration baseline on the source systems resulted in the increase in queries
Unexpected or unauthorized databased modification or destruction command(s) (e.g., DROP TABLE, ALTER TABLE)	• Network security team • Team responsible for network resource • Owner of the account that made the modification	1 hour	• Identify user or process responsible for command issuance • Identify other potential observables associated with nefarious actions on the system
Changes to vendor-supported baselines to include new accounts or access attempts by service or vendor accounts	• Support vendors • Network security team	48 business hours	• Validate if the change or account access correlates to known activity or vendor service window • Validate vendor access controls like time limits on interactive sessions
Theft of or unusual access to decoy documents	• Network security team • Team responsible for network resource	1 hour	Validate if activity is network noise or indicates potential malicious activity

ANALYZE INFORMATION AND CONTEXT COLLABORATIVELY

Documentation and Knowledge Management Process

Begin a documentation and knowledge management process in support of the investigation. Start with a determination of perceived actions in the triggering event. Did the change occur within:

- The physical domain (something involving telemetered quantities such as noise, voltage, current, frequency, or temperature, or the physical configuration of a piece of infrastructure); or
- The OT domain (something involving traffic or signals transiting a communication medium, or the logical configuration of a piece of infrastructure, or something involving memory/storage metrics); or

- Both the physical and the OT domains?

Given the determination of a physical or OT starting point, identify what expected corresponding perceptible observables would exist in the other domain and search for their presence or absence. For example, an unplanned file server reboot might produce digital footprints like logs and errors. It might correspond to data loss, unauthorized configuration changes, elevated user permissions, or account activity at unusual hours.

Technical Analysis

The order for which technical analysis should occur, or whether it is even necessary in the comprehension stage, depends on the situation; typically, it will inform many of the context-building questions outlined in the following section.

Technical analysis should first identify the scope of impacted information repositories and all peripheral systems and resources that might contain relevant data. Attacks using the T811 technique will likely require network traffic, host log, and file system analysis reports to properly comprehend the context and nature of the attack and determine how/if the triggering event is connected to other attack techniques. Network analysis may include flow analysis and packet inspection, depending on circumstance.

Vulnerability scans can be useful after a triggering event to determine whether a vulnerability has been introduced since the last scan. This can expose specific changes that introduced the vulnerability and can provide insight into the context of the triggering event.

Configuration files should be compared to a copy of a baseline configuration file to determine if improper or malicious configuration changes have occurred. If no baseline exists, suspect configuration files should be analyzed.

User and account permissions should be analyzed for unauthorized or improper changes. It is not uncommon for an attacker to escalate user privileges before accessing information repositories.

Context-Building Questions

A root cause analysis for suspected attacks using data from information repositories should begin with analysis of the repositories suspected to be involved in the breach. Due to the variety of services in industrial environments, analysis might need to first identify applications and systems that produce and consume data from information repositories. Consider the following questions:

- How was the attack identified? Context building should originate with either the system directly observed as being under attack or with the system reporting the highest severity of operational errors. The initial origin point of analysis might require an operator or analyst to leverage their knowledge and experience within the environment and intuition to determine the ideal starting point.

- What is "normal"? Determination of significant operational impact might involve discussions with operators. While operators might not be security experts, they do understand the physics of the environment and might also have a level of intuition as to what "normal" is within the environment.
- What was the attacker's goal? Determination of systems under attack might involve a review of host data or network data. Comparing alarm setting to previous backups can provide insight into attack motivation and/or desired impact.

Context building for data-gathering attacks on information repositories should uncover what data was affected and how. Because there are multiple information repositories within any system, it is important to use network and host data to isolate the portion of the network and specific host responsible for the traffic. This context might also assist in uncovering a wider intrusion into the environment.

Visual Information Modeling

Enumerate all the lines of questioning identified through this stage of the analysis of the anomaly. At this point, it is beneficial to begin a node and link diagram from the information documented in the knowledge management processes to help visualize relationships between observables; the CyOTE program colloquially refers to this observable linking diagram as a "worm diagram."

The triggering event is the first node, with all its related observables radially connected to it; include both confirmed observables and observables that were expected but not found, with some sort of visual discriminator between presence and absence (e.g., solid or dashed lines). Highlight the triggering event if the event relates to the implementation of a specific adversary technique. Include links emanating from the triggering event representing the as-yet-unanswered questions considered, as well as links and additional nodes for answered questions that confidently satisfy the extent of a particular line of inquiry. A notional example of this diagram for an investigation in progress is shown in Figure 4.

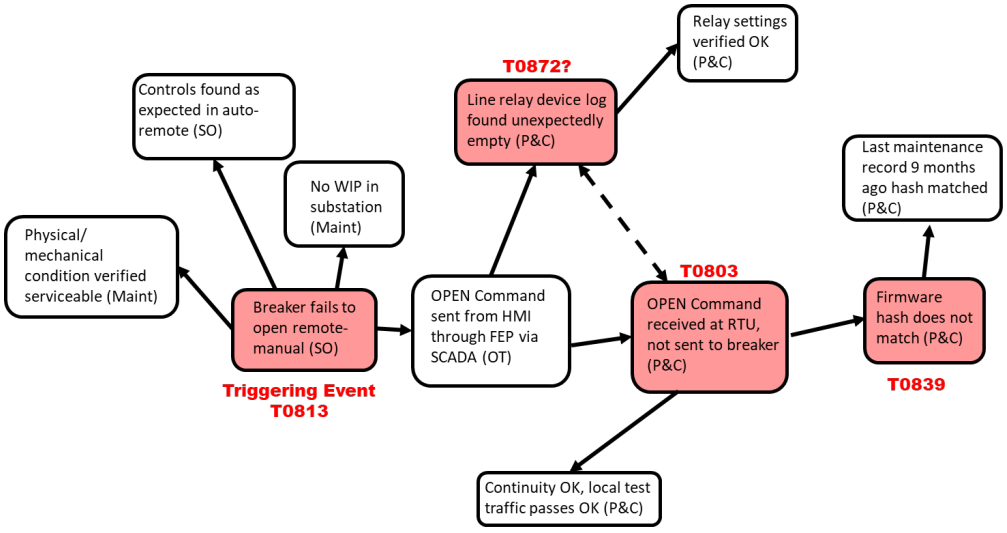

Figure 4: Example CyOTE Observables Link Diagram

INVESTIGATE POTENTIALLY RELATED ANOMALIES TO DATA FROM INFORMATION REPOSITORIES

After an anomaly has been comprehended sufficiently and mapped to a technique, repeat the steps above, starting from each questioning line resulting from the initial anomaly analysis. The importance of recording and organizing the information discovered in the comprehension process and visualizing it through a node and link diagram becomes exponentially more important as the first perceived anomaly expands into a web of postulated, confirmed, and denied relationships.

DECISION: REMEDIATING THE CAUSE AND EFFECTS OF A TRIGGERING EVENT

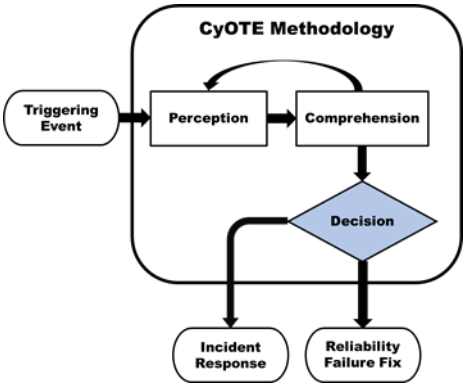

Figure 5: CyOTE Methodology - Decision Step

The last step of the CyOTE methodology is the Decision step (Figure 5). In this step, the AOO makes a risk-informed business decision on how to proceed based on the information gathered in the previous steps. The level of comprehension and detail needed to make the decision will vary from company to company and may be related to resource availability. The length and consistency of a discovered and comprehended "worm diagram" representing a prospective kill chain fragment needed to decide to proceed with incident response will also vary based on a company's risk tolerance.

INCIDENT RESPONSE

In situations where there is sufficient belief that the anomalies perceived and comprehended indicate possible malicious cyber activity, the appropriate organizational action is to initiate a cybersecurity incident response process according to established policy and procedures. The information and context developed through the application of the CyOTE methodology will be helpful to incident handlers for developing and implementing appropriate mitigating actions. Consult with your organization's incident response procedures for the next steps.

CORRECTIVE MAINTENANCE

When analysis fails to establish a reasonable indication of malicious cyber activity, the next step is to determine if the anomalies indicate that a non-malicious failure is occurring. One appropriate action involves resolving any deficiencies discovered through corrective maintenance and work management processes according to organizational policy and procedures. It is worthwhile to maintain records of these situations for future reference and comparison to subsequent anomalies. Consult with your organization's engineers, operators, and other parties responsible for maintenance in your environment. For vendor-supported devices, coordinate with external parties when necessary.

IMPROVING SECURITY CAPABILITIES

Prevention and preparation include proactive measures to reduce the risk or likelihood of an attack or enable an organization to respond more rapidly to a situation. Preventive actions may raise the cost of an attack, which could deter the attacker from an act or eliminate the possibility that specific attack paths would be used altogether. AOOs should develop and deploy data collection management strategies to prepare for the collection of data needed to detect this technique and properly analyze potentially anomalous triggers.

CONTROL MATRIX

The control matrix (Table 4) is a technical capability or system designed to prevent or sabotage specific attack techniques used by malicious actors. The control matrix assists with the identification of detection capability improvement areas.

Table 4: Control Matrix

Control	Matrix	Relevance
Multi-Factor Authentication	• MITRE D3FEND: D3-MFA[11] • NIST 800-53: IA-5[12] • NIST 800-82: 6.2.7.3[13] • NIST 800-82: 6.2.7.4[14] • NIST 800-82: 6.2.7.5[15]	Multi-factor authentication should be used on all supported information repositories to raise the level of rigor applied to user validation. • Identify existing single-factor authentication systems connected to information repositories. This should include physical access and authentication systems in addition to digital systems. • Develop policies to ensure MFA implementations are in place for existing technology and future acquisitions. • Consider MFA capabilities when acquiring 3rd party software solutions like directory services, file servers, or OT interfaces. • Consider that internal systems used as an additional authentication factor might already be compromised. For example, internal email systems that have already been unknowingly compromised would assist attackers in the case of internal email verification.
Disk Encryption	• MITRE D3FEND: D3-DENCR[16] • MITRE ICS ATT&CK: M0941[17]	Disk Encryption is used to protect information repositories from unauthorized access to readable data. Digital storage devices and/or their partitions should be encrypted to prevent adversaries from having cleartext access to a file system. • Identify existing digital storage devices and file systems not utilizing data encryption. This includes archived file systems and digital storage devices not in use and/or in storage. • Ensure OPSEC policy and training includes encryption requirements, including secure key handling. • Data should be securely wiped before disposal even if it is encrypted. • Consider that file systems currently in use by OT are decrypted and vulnerable to clear-text access by adversaries.
RF Shielding	MITRE D3FEND: D3-RFS[18]	RF Shielding prevents insider adversaries from accessing information repositories through side-channels. Digital devices emit unique radio-frequency signals corresponding to

[11] MITRE, Multi-Factor Authentication, 2021. Available online: https://d3fend.mitre.org/technique/d3f:Multi-factorAuthentication/.
[12] NIST, Authenticator Management, 2021. Available online: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=IA-5.
[13] NIST, NIST Special Publication 800-82: Revision 2, 2015. Available online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf.
[14] Ibid.
[15] Ibid.
[16] MITRE, Disk Encryption, 2021. Available online: https://d3fend.mitre.org/technique/d3f:DiskEncryption.
[17] MITRE, Encrypt Sensitive Information, 2021. Available online: https://collaborate.mitre.org/attackics/index.php/Mitigation/M0941.
[18] MITRE, RF Shielding, 2021. Available online: https://d3fend.mitre.org/technique/d3f:RFShielding.

Control	Matrix	Relevance
		their specific digital operations which can be analyzed to reveal secret data, including credentials or other vectors for information repository access. ● Ensure relevant components supplied by 3rd parties implement effective shielding. ● Consider implementing policies regarding standards for RF shielding for both currently operating equipment and future acquisitions. ● Require the use of a device, display, and cable shielding designed to reduce RF interference.
Software Update	MITRE D3FEND: D3-SU[19]	Software Updates reduce the attack surface of OT software. For T811, it is necessary to update any software or application which authenticates or facilitates user access to information repositories. ● Identify and document all software, applications, and their versions. Additionally, keep an updated log/history for documented software. ● Implement systems to alert relevant personnel to newly available software versions. ● Implement systems to alert relevant personnel to any CVEs, vulnerabilities, or recent exploits to current or previous software versions used to authenticate or facilitate access to information repositories. ● Consider that a newly exploited vulnerability in an older software version may not yet be patched in an up-to-date version. ● Develop a policy regarding software update standards and requirements. Consider employees, contractors, remote access applications, and software acquisition. ● Develop guidelines to employ authentication mechanisms (like hashing) for software update sources.
File Hashing	MITRE D3FEND: D3-FH[20]	File Hashing is a file analysis security control used to detect known malware signatures within files using a list of hashes to compare against. Adversaries can target and infect otherwise legitimate files with malware to be uploaded to information repositories for the purposes of exfiltrating data, implanting a backdoor, or deploying ransomware. ● Implement a file hashing capability to continuously monitor files within information repositories for malware signatures. ● Hash lists are compiled and distributed by commercial vendors. ● Consider performance costs for hashing large amounts of

[19] MITRE, Software Update, 2021. Available online: https://d3fend.mitre.org/technique/d3f:SoftwareUpdate.
[20] MITRE, File Hashing, 2021. Available online: https://d3fend.mitre.org/technique/d3f:FileHashing.

Control	Matrix	Relevance
		files. ● Ensure hash lists are up to date. ● File hashing will not detect malware signatures that are not included in the list of hashes. ● Obfuscated malware can avoid detection from simple hashing techniques.
File Content Rules	MITRE D3FEND: D3-FCR[21]	File Content Rules is a file analysis security control used to detect both generic and targeted malware by employing a pattern matching rule language (DSL: Domain Specific Language) in file analysis. Adversaries can target and infect otherwise legitimate files with malware to be uploaded to information repositories for the purposes of exfiltrating data, implanting a backdoor, or deploying ransomware. ● Identify gaps or weaknesses in existing file analysis and malware detection capabilities. ● Implement rule-based detection security controls to compliment other existing file analysis security controls. ● Consider the difficulty and resources required to write new rules. ● Consider the tradeoff between trigger/anomaly fidelity and computational load. ● Consider adversarial attempts to design malware in such a way as to exploit or defeat the scanning engines. ● Obfuscated malware can avoid detection from rule-based file analysis security controls.
Administrative Network Activity Analysis	MITRE D3FEND: D3-ANAA[22]	Administrative Network Activity Analysis is a network traffic analysis security control that monitors and analyzes remote administrative network protocol activity against a baseline. Protocols used by administrators to manage systems and services related to information repositories can be abused by attackers who have gained unauthorized access to administrator accounts. ● Monitor and log existing administrator network traffic metadata over a period of time to establish baseline behavior. ● Implement a capability which analyzes baseline administrator behavior and compares it to current activity to detect anomalous activity. ● Establish an alerting mechanism to make relevant personnel aware of anomalous events. ● Increase trigger fidelity by developing a system to update the baseline with activity causing false alarms. ● Establish policy for authorized administrator activity on

[21] MITRE, File Content Rules, 2021. Available online: https://d3fend.mitre.org/technique/d3f:FileContentRules.
[22] MITRE, Administrative Network Activity Analysis, 2021. Available online: https://d3fend.mitre.org/technique/d3f:AdministrativeNetworkActivityAnalysis.

Control	Matrix	Relevance
		information repositories. Administrator accounts typically do not need to read/write data within information repositories other than that which is required for specific administration tasks.
Client-Server Payload Profiling	MITRE D3FEND: D3-CSPP[23]	Client-Server Payload Profiling is a network traffic analysis security control that compares client-server request and response payloads to a baseline to detect anomalous activity between clients and information repositories. ● Identify systems containing information repositories and monitor request and response payloads over the network to develop baseline behavior profiles for client-server interactions. ● Consider that there are many factors that can reduce anomaly/trigger fidelity, including: sharded user accounts, changing user behavior, inconsistent work schedules, and user login/entry points, among others. ● This security control can generate a lot of "noise" or false signals, but depending on the implementation, there are some trigger or alerting metrics that can be employed to identify obvious malicious behavior. An example would be thresholds for very large and continuous payloads, indicating large data exfiltration attempts.
Per Host Download-Upload Ratio Analysis	MITRE D3FEND: D3-PHDURA[24]	Per Host Download-Upload Ratio Analysis is a network traffic analysis security control that compares baseline data upload-to-download ratios for each host to identify unusual activity. This security control can help identify some of the more obvious data exfiltration attempts. ● Monitor and log existing metadata push vs. pull ratios for each host over a period of time to establish baseline ratio metrics for typical host upload-to-download behavior. ● Consider that large packet captures require significant computing and storage resources. ● Consider that baselines need to be established for new hosts. ● This security control can generate a lot of "noise" or false signals, but depending on the implementation, there are some trigger or alerting metrics that can be employed to identify obvious malicious behavior. An example would be unusually consistent push vs. pull ratios, potentially indicating automated data exfiltration.
Local Account Monitoring	● MITRE ICS ATT&CK: M0947[25] ● MITRE ICS ATT&CK:	Local Account Monitoring is a platform monitoring security control used to detect unauthorized interaction with

[23] MITRE, Client-Server Payload Profiling, 2021. Available online: https://d3fend.mitre.org/technique/d3f:Client-serverPayloadProfiling.
[24] MITRE, Per Host Download-Upload Ratio Analysis, 2021. Available online: https://d3fend.mitre.org/technique/d3f:PerHostDownload-UploadRatioAnalysis.
[25] MITRE, Audit, 2021. Available online: https://collaborate.mitre.org/attackics/index.php/Mitigation/M0947.

Control	Matrix	Relevance
	M0926[26] • MITRE ICS ATT&CK: M0918[27] • MITRE D3FEND: D3-LAM[28] • NIST 800-53: SI-7(8)[29] • NIST 800-53: AC-2[30] • NIST 800-82: 6.2.3[31] • NIST 800-82: 6.2.17[32]	information repositories from local user accounts. • Establish policy and guidelines regarding user access to information repositories. These rules will serve as alert metrics for unauthorized activity. • Utilize role-based access policies and restrictions in the development of monitoring and alert criteria. • Ensure there is one user per account wherever possible to increase trigger fidelity and simplify analysis upon alerting activity. • Monitor account activity for a period of time to establish baseline activity to compare against in the case of alerting activity. • Consider that this security control can cause a large amount of false positives without well-developed baselines, user account policies, and thresholds.
Input Device Analysis	MITRE D3FEND: D3-IDA[33]	Input Device Analysis is a platform monitoring security control used to prevent data exfiltration from information repositories to external media. • Implement capability to filter input device commands or disable them entirely. • Closely monitor behavior from input devices. • Determine metrics to monitor like date/time, system processing the input, info and properties of the input device, source (local vs. remote), active user, etc. • Responses to unusual activity can include disabling the device, session, or credentials. • External devices can be used as an exfiltration device. Monitor for traffic being routed through a new input device correlated with a similar volume of traffic exiting the previously existing interface. • Consider a policy prohibiting all USB device usage.
Database Query String Analysis	• MITRE D3FEND: D3-DQSA[34]	Database Query String Analysis is a process analysis security control used to detect unauthorized accesses to information repositories by malicious database queries.

[26] MITRE, Privileged Account Management, 2021. Available online: https://collaborate.mitre.org/attackics/index.php/Mitigation/M0926.
[27] MITRE, User Account Management, 2021. Available online: https://collaborate.mitre.org/attackics/index.php/Mitigation/M0918.
[28] MITRE, Local Account Monitoring, 2021. Available online: https://d3fend.mitre.org/technique/d3f:LocalAccountMonitoring.
[29] NIST Risk Management Framework; Software, Firmware, and Information Integrity; "SI-7(8): Auditing Capability for Significant Events;" 2021. Available online: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=SI-7.
[30] NIST Risk Management Framework, "AC-2: Account Management," 2021. Available online: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=AC-2.
[31] NIST Special Publication 800-82, "Guide to Industrial Control Systems," Revision 2, 2015. Available online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf.
[32] Ibid.
[33] MITRE, Input Device Analysis, 2021. Available online: https://d3fend.mitre.org/technique/d3f:InputDeviceAnalysis.
[34] MITRE, Database Query String Analysis, 2021. Available online: https://d3fend.mitre.org/technique/d3f:DatabaseQueryStringAnalysis.

Control	Matrix	Relevance
	• NIST 800-53: SI-7[35] • NIST-800-82: 6.2.17[36]	• Implement a capability to monitor and alert relevant personnel upon predetermined SQL database queries or query patterns. • Consider user roles in determining alerting activity. • Monitor user input for SQL injection attacks. • Consider that improperly implemented input sanitization can introduce vulnerabilities, result in undesired changes to data, and break functionality of systems relying on database data for operations.
User Data Transfer Analysis	• MITRE D3FEND: D3-UDTA[37] • NIST 800-53: SI-4[38]	User Data Transfer Analysis is a user behavior analysis security control that can detect data exfiltration from information repositories by analyzing the amount of data transferred by a user. • Implement a network traffic and application logging capability for user data transfer metrics. • Establish thresholds and baseline metrics for typical data transfer behavior and implement an alarm mechanism for activity that needs to be reviewed. • Consider attack techniques employed to blend in with typical user data transfer activity. • This security control can have a high rate of false positives without effective thresholds and baselines.
Mandatory Access Control	• MITRE D3FEND: D3-MAC[39] • MITRE ICS ATT&CK: M0922[40]	Mandatory Access Control by file path level controls is a kernel-level execution isolation security control used to control access to information repositories by pathname level access policies. • Access policy must be implemented before developing or acquiring a capability for this security control. • Some implementations can be complex and difficult to maintain over time.
Inbound Traffic Filtering	MITRE D3FEND: D3-ITF[41]	Inbound Traffic Filtering is a network isolation security control that restricts network traffic originating from untrusted networks toward devices or systems containing information repositories. • Identify networks containing devices and systems with

[35] NIST Risk Management Framework, "SI-7: Software, Firmware, and Information Integrity," 2021. Available online: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=SI-7.
[36] NIST Special Publication 800-82, "Guide to Industrial Control Systems," Revision 2, 2015. Available online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf.
[37] MITRE, User Data Transfer Analysis, 2021. Available online: https://d3fend.mitre.org/technique/d3f:UserDataTransferAnalysis.
[38] NIST Risk Management Framework, "SI-4: Information System Monitoring," 2021. Available online: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=SI-4.
[39] MITRE, Mandatory Access Control, 2021. Available online: https://d3fend.mitre.org/technique/d3f:MandatoryAccessControl.
[40] MITRE, Restrict File and Directory Permissions, 2021. Available online: https://collaborate.mitre.org/attackics/index.php/Mitigation/M0922.
[41] MITRE, Inbound Traffic Filtering, 2021. Available online: https://d3fend.mitre.org/technique/d3f:InboundTrafficFiltering.

Control	Matrix	Relevance
		information repositories and develop a capability to filter incoming network traffic. ● Implement filters considering the OT environment needs and requirements. ● Implementation might block certain IPs and spoofed addresses, or specific ports, services, or IP ranges. ● Consider dynamic inbound filtering. ● Consider that protocols using non-standard ports can circumvent detection.
Outbound Traffic Filtering	MITRE D3FEND: D3-OTF[42]	Outbound Traffic Filtering is a network isolation security control that helps isolate networks within the OT environment and prevents network traffic which may contain data from information repositories being exfiltrated to an untrusted network. ● Identify networks containing devices and systems with information repositories and develop a capability to filter outgoing traffic to prevent data leaks or exfiltration. ● Establish clear network boundaries and digital policies preventing user accounts from sending data from information repositories to untrusted networks or hosts. ● Consider challenges in designing filters due to dynamic IP assignment. ● Connections using non-standard transport layer ports can circumvent detection.
Connected Honeynet	MITRE D3FEND: CHN[43]	Connected Honeynet is a decoy environment security control that uses decoy systems, services, or environments connected to networks containing information repositories, and simulates or emulates some functionality without exposing full access to the system. This security control can act as a sensor by alerting relevant personnel to malicious interactions with information repositories before an attack can progress to legitimate systems. ● Identify known vulnerabilities or system weaknesses that might attract malicious actors. ● Design a decoy system, service, or environment (honeynet) that an attacker would perceive as a low effort vector to gain access to information repositories. ● The decoy honeynet should be connected to the OT, IT, and/or enterprise environment in such a way that it does not expose full access to the environment but looks as if it is legitimate to malicious actors. ● Consider that improperly designed honeynets can increase the attack surface and reduce the overall security of the system.

[42] MITRE, Outbound Traffic Filtering, 2021. Available online: https://d3fend.mitre.org/technique/d3f:OutboundTrafficFiltering.
[43] MITRE, Connected Honeynet, 2021. Available online: https://d3fend.mitre.org/technique/d3f:ConnectedHoneynet.

Control	Matrix	Relevance
Standalone Honeynet	MITRE D3FEND: D3-SHN[44]	Standalone Honeynet is a decoy environment security control that uses decoy systems, services, or environments that are logically disconnected from all legitimate OT, IT and/or enterprise systems. ● Standalone honeynets may share the same address space as legitimate systems but must not interact with those systems. ● Consider the tradeoff of a standalone honeynet being lower risk to deploy but at the cost of realism. Making the honeynet look realistic can take significant effort and resources.
Decoy File	● MITRE D3FEND: D3-DF[45] ● MITRE D3FEND: D3-DNR[46] ● MITRE D3FEND: D3-DUC[47] ● NIST 800-53: SC-26[48] ● NIST-800-82: 6.2.17.2[49]	Decoy File is a decoy object security control that uses a monitored file to attract interaction from malicious actors. This decoy file effectively acts as a sensor to detect malicious interaction with information repositories and alert relevant personnel before an attack progresses. ● Decoy files may be made available as documents containing information relating to the ICS, or as network resources, configuration files, email attachments, or any other file that attackers might perceive as vectors to access information repositories. ● Implement a capability to monitor decoy files and alert relevant personnel in the case of interaction with these files. ● File properties should be modified in such a way that an attacker perceives it as legitimate.
Decoy Network Resource	MITRE D3FEND: D3-DNR[50]	Decoy Network Resource is a decoy object security control that uses monitored network resources to attract interaction from malicious actors. The decoy network resource is deployed to web app servers or network-based file sharing services and effectively acts as a sensor to detect malicious interaction with network resources related to information repositories and alert relevant personnel before an attack progresses. ● Identify strategic locations to deploy decoy network resources and the type of network resources to deploy.

[44] MITRE, Standalone Honeynet, 2021. Available online: https://d3fend.mitre.org/technique/d3f:StandaloneHoneynet.
[45] MITRE, Decoy File, 2021. Available online: https://d3fend.mitre.org/technique/d3f:DecoyFile.
[46] MITRE, Decoy Network Resource, 2021. Available online: https://d3fend.mitre.org/technique/d3f:DecoyNetworkResource.
[47] MITRE, Decoy User Credential, 2021. Available online: https://d3fend.mitre.org/technique/d3f:DecoyUserCredential.
[48] NIST Risk Management Framework, SP 800-53 Rev. 4.0, "SC-26: Honeypots," 2021. Available online: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=SC-26.
[49] NIST, NIST Special Publication 800-82, "Guide to Industrial Control Systems," Revision 2, 2015. Available online: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf.
[50] MITRE, Decoy Network Resource, 2021. Available online: https://d3fend.mitre.org/technique/d3f:DecoyNetworkResource.

Control	Matrix	Relevance
		● Implement a capability to monitor decoy network resources and alert relevant personnel in the case of interaction with these resources. ● Consider the possibility of DoS attacks.
Account Locking	MITRE D3FEND: D3-AL[51]	Account Locking is a credential eviction security control that disables or limits some functionality of involved user accounts upon alerting events. This provides relevant personnel the time to comprehend the nature and effects of an attack and inform decisions. ● This security control requires some type of account management capability with the ability to set rules and policies for accounts and enable/disable accounts. ● Consider accounts associated with mission critical operations and the consequences of disabling such an account. ● Consider role-based vs. attribute-based systems and centralized vs. local cache-based implementations. ● Weigh the severity level of anomalous events and triggers against the consequences of locking an account. ● Consider that improperly designed account locking implementations could increase the attack surface and provide a vector for attackers to disrupt critical operations.
Authentication Cache Invalidation	MITRE D3FEND: D3-ANCI[52]	Authentication Cache Invalidation is a credential eviction security control that removes user account authentication data from a cache to deny further account access to certain resources. This provides relevant personnel the time to analyze anomalous activity while allowing the user account access to limited functionality. ● This security control requires some type of account management capability with the ability to set rules and policies for accounts and enable/disable accounts. ● Consider accounts associated with mission critical operations and the consequences of disabling such an account. ● Consider that credentials can be cached on a remote server in addition to being cached locally. ● Weigh the severity level of anomalous events and triggers against the consequences of restricting user interaction with certain services. ● Consider that improperly designed credential eviction implementations could increase the attack surface and provide a vector for attackers to disrupt critical operations.

[51] MITRE, Account Locking, 2021. Available online: https://d3fend.mitre.org/technique/d3f:AccountLocking.
[52] MITRE, Authentication Cache Invalidation, 2021. Available online: https://d3fend.mitre.org/technique/d3f:AuthenticationCacheInvalidation.

Control	Matrix	Relevance
Connection Attempt Analysis	• MITRE D3FEND: D3-CAA[53] • NIST 800-53: SI-4[54] • NIST 800-82: 6.2.17.2[55]	Connection Attempt Analysis is a network traffic analysis security control used to monitor connection attempts to common ports associated with systems containing information repositories. • Identify network and system boundaries which will serve as criteria for connection attempt analysis. • Implement a capability to monitor connection attempts to systems with information repositories. • Connections from untrusted networks or systems should be blocked and relevant personnel should be alerted. • Collect and save collection data using a rotating log implementation based on the needs and capabilities of the organization. This can help inform analysis and serve in developing baselines.

TESTING, DEPLOYING, AND REFINING ALERTING PARAMETERS FOR DATA FROM INFORMATION REPOSITORIES

The parameters and established triggers from previous analysis lead to the continual testing and gradual deployment of the capability and evaluation of alerts. The resulting output functions and considerations are realized during this process, the products of which are sent to output destination(s) in selected format(s).

1. Validate triggers and alerts
    a. Ensure the capability does not conflict with existing monitoring functionality
    b. Ensure the capability does not adversely impact the existing environment
    c. Test alerting functions
        i. Use synthetic data (e.g., PCAPs)
        ii. If the test fails, re-evaluate the steps taken iteratively (line by line)
        iii. If successful, enact a graduated deployment schedule and retest for each iteration
    d. Consider communication criteria for multiple locations and information consolidation during graduated deployment
2. Identify output destination(s) (SIEM, Splunk, Gravwell, Elk)
    a. Identify output format(s) (STIX, Syslog, JSON, CSV)
    b. Define actionable data requirements, processes, and responses
        i. Logging
        ii. Alert content

[53] MITRE, Connection Attempt Analysis, 2021. Available online: https://d3fend.mitre.org/technique/d3f:ConnectionAttemptAnalysis.
[54] NIST Risk Management Framework, "SI-4: Information System Monitoring," 2021: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=SI-4.
[55] NIST, NIST Special Publication 800-82, "Guide to Industrial Control Systems," Revision 2, 2015. Available online: https://csrc.nist.gov/Projects/risk-management/sp800-53-controls/release-search#!/control?version=4.0&number=SI-4.

iii. Alert response(s) (local or SOC)

3. Identify what information to log (long-term/short-term)
    a. The aggregation of different log types may assist in identifying potentially anomalous behaviors within OT environments

The overall output of this process may result in one of the following: script, application, Yara rule, SIEM plugin, or another tool capable of identifying and alerting on the use of the Data from Information Repositories technique within OT environments.

CONCLUSION

By employing the strategies in each step of the CyOTE methodology outlined in this Recipe, AOOs will be able to develop a capability to read and analyze their OT traffic and determine whether the Information from Data Repositories technique is being used as part of a malicious cyber-attack.

Through the Perception step, the AOO identifies observables and anomalies indicative of the Information from Data Repositories technique. Determining the associated observables and potential data sources for these anomalies will give the AOO the basis they need to begin comprehending the context in which a triggering event involving the Information from Data Repositories technique came to be. Performance degradation when exfiltrating data; suspicious network traffic to a data repository; unexpected file access and file storage in a data historian; and high-risk queries to a data repository are all potential observables that could indicate the use of the Information from Data Repositories technique. Anomalies tied to these observables could include unusual file or folder collections on a data repository, unexpected crashes with critical historian services, unauthorized file metadata change or other changes to baseline code, or attempted access of decoy documents or information.

The Comprehension step is used to collect information around observables and anomalies via the data sources identified in the previous step. AOOs should leverage various business organizations and areas of the OT environment for information for identifying and analyzing anomalies related to the Information from Data Repositories technique. This will allow them to more quickly identify triggering events using the Information from Data Repositories technique, enabling the decision making in the Decision step.

AOOs use the Decision step to determine a course of action based on an event that has occurred. With the information gathered, the AOO will be able to determine whether anomalies perceived in a data repository is indicative of an adversary's presence in the network (thus initiating incident response procedures) or if there is no evidence of malicious activity associated with the anomalies (thus initiating corrective maintenance procedures).

Additional assistance regarding general sensor placement and capability development is available through DOE; contact CyOTE.Program@hq.doe.gov for more information. AOOs can also refer to the CyOTE methodology for more information on CyOTE's approach to identifying anomalies in an OT environment, which, when perceived, initiates investigation and analysis to comprehend the anomaly.

APPENDIX A: DATASETS TO SUPPORT TRIGGERING EVENT ANALYSIS FOR T811: DATA FROM INFORMATION REPOSITORIES

Table 5: Datasets to Assist with Analyzing Triggering Events

Dataset	Example Tools	Who Can Assist	Relevance
Netflow and Packet Data	• Wireshark/TShark • Commercial Passive Network Monitoring Tools (Claroty, Dragos, Nozomi, SilentDefense) • Zeek • NetworkMiner • Snort • Suricata • Security Onion	• Network Security Team • IT or OT System Admins	Netflow and packet data assists with identification of systems communicating and possibly detailed communication details
Device & System Logs	• Sysinternals Sysmon • Sysinternals PsLogList • EvtxToElk • Python-evtx • Osquery	• Network Security Team • IT or OT System Admins	Device & system log data assists with identification of systems communicating and possibly detailed communication details
Device & System Configuration Files and Change History	Sysinternals Suite	• Network Security Team • IT or OT System Admins	Device & system configuration files and change history assists with identification of systems communicating and possibly detailed communication details
Account administration data like permission settings, account logs, onboarding information	Sysinternals Suite	• Network Security Team • IT or OT System Admins	Permission settings, account logs and onboarding information for accounts can assist with discovering other associated behaviors with the trigger event in question
Device or System Maintenance Documentation/Logs	Sysinternals Suite	• Network Security Team • IT or OT System Admins	Device or system maintenance document and logs assists with identification of systems communicating and possibly detailed communication details

Dataset	Example Tools	Who Can Assist	Relevance
Physical access logs and security monitoring data like CCTV output	Application Specific	Physical Security Team	Physical security logs and CCTV adds another factor of validation to assist with validation of true source
System engineering documents like network layouts and other schematics or diagrams	Diagram Specific	● Network Security Team ● IT or OT System Admins ● OEMs/Third-Party Vendors	Environment documentation assists with identification of other logging sources or impacted systems
Lists of software, apps, hardware, devices, or other relevant systems and their respective manufacturers	Asset Inventory Tools (Claroty, Dragos, Nozomi, SilentDefense)	● Network Security Team ● IT or OT System Admins	Software and hardware lists assist with identification of other impacted systems as well as other potential log resources to validate a trigger event
Any other data relevant to the investigation	Various	● Network Security Team ● IT or OT System Admins ● OEMs/Third-Party Vendors	Other data sources might contain information specific to a given trigger event

Click for More Information	CyOTE Program || Fact Sheet || CyOTE.Program@hq.doe.gov
DOE Senior Technical Advisor	Edward Rhyne || Edward.Rhyne@hq.doe.gov || 202-586-3557