INL/EXT-21-61519-Revision-0

Neighborhood Keeper Program Review

January 2021

Jeffrey L Hahn, Christopher M Spirito, Justin N Cox, Lawrence R Wellman, Timothy Conway

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/EXT-21-61519-Revision-0

Neighborhood Keeper Program Review

Jeffrey L Hahn, Christopher M Spirito, Justin N Cox, Lawrence R Wellman, Timothy Conway

January 2021

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract 20SP860

SPP No. 20SP860
Revision 1

Neighborhood Keeper Program Review

January 13, 2021

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory
operated by Batelle Energy Alliance, LLC

SPP No. 20SP860
Revision 1

Neighborhood Keeper Program Review

January 13, 2021

Idaho National Laboratory
Cybercore Integration Center
National and Homeland Security
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for
Dragos, Inc.
Prime Contractor of US DOE Funding Opportunity
No. DE-FOA-00001755
SPP No. 20SP860

Page intentionally left blank

CONTENTS

Table of Abbreviations..... v
INTRODUCTION..... 1
DISCUSSION..... 1
Recommended Detections or Combinations of Detections..... 1
Additional Analysis that Could Benefit the Electricity Subsector.....2
Recommended Modification to Neighborhood Keeper Reports.....3
Data Sharing Lessons Learned Through INL's Experience With Utilities..... 6
Recommendations that Could Benefit the Electricity Subsector..... 7
Appendix A Use of Artificial Intelligence and Machine Learning.....9
Appendix B National Laboratory Artificial Intelligence and Machine Learning Projects..... 15

Page intentionally left blank

Table of Abbreviations

ADS	Anomaly Detection System
AI	Artificial Intelligence
AICS	Automatic Intelligent Cyber Sensor
BES	Bulk Electric System
CDA	Critical Digital Assets
CEDS	Cybersecurity of Energy Delivery Systems
CIP	Critical Infrastructure Protection
CPAD	Cyber-Physical Attacks Detection
CSI	Cyber System Information
DCS	Distributed Control System
DNP3	Distributed Network Protocol 3
EACMS	Electronic Access Control or Monitoring Systems
ELK	Elasticsearch, Logstash, and Kibana
EMS	Energy Management System
HMI	Human Machine Interface
ICS	Industrial Control System
INL	Idaho National Laboratory
IP	Internet Protocol
IREST	ICS Resilient Security Technology
IT	Information Technology
JSON	JavaScript Object Notation
ML	Machine Learning
NERC	North American Electric Reliability Corporation
OSI	Open System Interconnection
OT	Operational Technology
OUO	Official Use Only
PCAP	Packet Capture
PCII	Protected Critical Infrastructure Information
SCADA	Supervisory Control and Data Acquisition
SME	Subject Matter Experts
SOC	Security Operations Center
SSEP	Safety, Security, and Emergency Preparedness
STIG	Structured Threat Intelligence Graph
STIX	Structured Threat Information eXpression
SVM	Support Vector Machine
TTP	Tactics, Techniques, and Procedures

Page intentionally left blank

Neighborhood Keeper Program Review

INTRODUCTION

Dragos Inc., the Prime Contractor, under U.S. Department of Energy Funding Opportunity Number DE-FOA-00001775: Industry Partnerships for Cybersecurity of Energy Delivery Systems (CEDS) Research, Development and Demonstration entitled "Neighborhood Keeper," has requested that Idaho National Laboratory (INL) provide technical expertise through Strategic Program Partnership No. 20SP860.

Dragos has provided INL access to their Neighborhood Keeper platform, which contains simulated data, sample reports given to utilities, and access to anonymized program data that is sent from utilities to the cloud. Dragos has specifically asked INL to review the available data and respond to the following questions:

- What detections or combinations of detections are most useful?
- What additional analysis would benefit the electricity subsector?
- How could Neighborhood Keeper reporting be modified to benefit the electricity subsector?
- Based on INL's experience with utilities, are there lessons learned about data sharing that could be provided?
- Are there recommendations that could benefit the electricity subsector?

DISCUSSION

Recommended Detections or Combinations of Detections

The Neighborhood Keeper program detections can be thought of in three buckets: 1) detection perspective; 2) Industrial Control System (ICS) activity context; and 3) programmatic situational awareness. The desire to improve capabilities in these three areas is not new. Previous efforts to gain increased understanding and visibility into adversary activities within critical infrastructure environments have always faced limitations to success. These limitations have historically included:

- Constrained or partial visibility into operational technology (OT) networks due to infrastructure constraints.
- Information security concerns impeding who can see what types of information across an organization and what can be shared with others.
- Isolated projects within an organization that may have identified adversary activities, suspicious events, or important lessons learned.

For these reasons and numerous others, industry has been working to assess adversary capabilities through a keyhole, rather than through a deeper collection and broader field of vision.

Detection Perspective

With an engineered operational process that is under control, sensor placement is of utmost importance. This allows the process to be evaluated and acted upon to maintain desired outcomes. For example, in a temperature control process, placing a thermocouple sensor outside of the process or directly in front of a fan blower would result in a sensor that is reading a continuous value that is of no use in controlling the process. This could also result in input values with high amounts of non-representative noise that cause the process to operate inefficiently. In a similar way, if we wish to defend a process environment from a

cybersecurity perspective, it is important we position network sensors appropriately to collect meaningful and actionable information. With a goal of defending the ICS environments across national critical infrastructure, consider the value of placing a network sensor at the business environment edge (similar to placing a thermocouple directly in front of a fan blower) or placing a network sensor within the business environment (similar to placing a thermocouple outside of the temperature-controlled area). The sensor would collect information with non-ICS context. While these network sensor locations and points of visibility may be important for other defender objectives and goals, they are not as effective in defending the process network.

Instead, defenders would achieve better results by implementing ICS-aware sensors within the OT networks. Network sensors positioned within the process environment provide much better sensor visibility and data genesis for the Neighborhood Keeper program. ICS-aware sensor placement is essential for the Neighborhood Keeper program and needs to be pursued across program participants during implementation.

Industrial Control System (ICS) Activity Context

When sensors are placed appropriately within the process environment, the Neighborhood Keeper program has the ability to begin to understand ICS-specific communications, unique data sets within a multi-vendor process environment, and methods in which activity groups work within the targeted OT networks. This visibility and cyber operator interface is available through the local Dragos platform solution. This element of the Neighborhood Keeper program may be one of the strongest components of the value proposition for participants. There is great value in the anonymized detections provided by the participants and the resulting general reports. This essential element of the program allows for continuous learning and improvement across the program as detections can be fully evaluated by the participants. Informed by program detections, or internal capabilities, participants can leverage the Neighborhood Keeper platform to investigate alerts of interest, leverage playbook-driven, consistent responses, and potentially identify novel adversary actions.

Programmatic Situational Awareness

While sensor placement and data genesis are foundational to the value of the program, an interesting aspect of the Neighborhood Keeper program is the wide area view that it provides across all participant environments. The program provides the community of cyber operators a supervisory control and data acquisition (SCADA) situational awareness equivalent of detections occurring across participating organizations. This will provide participants with visibility to adversary activities beyond their own organization. It also offers participants the ability to take these indicators and begin detection and hunt activities where sensors may not exist within their own OT networks. Neighborhood Keeper program stakeholders will see great value in this wide area view capability as participants can identify a coordinated activity or a broad campaign targeting multiple critical infrastructure entities.

Attempting to evaluate which detections or combinations of detections are most valuable is a bit of an impossibility without understanding each participant's operational environment. Proper evaluation will depend on many variables, such as potential safety impacts, current system conditions, unique customer load serviced, and asset variables under the control of the operator. In this way, individual participants and other program stakeholders will evaluate, identify, and expand upon the true value of the detections. While Neighborhood Keeper is in its early stages, the uniqueness of the program is in the combination of the data genesis, ICS-specific context, and wide area view capabilities it provides. These benefits are expected to grow with additional participants.

Dragos has implemented a unique approach to data sanitization that reduces sensitivities around regulatory data protection, information access, and information sharing.

Additional Analysis that Could Benefit the Electricity Subsector

Data Set Time Horizons

Electric system operators often consume and act upon operational data in three different time horizons: ahead of time (planning), real time (operational), and after the fact (analysis). Cyber operators will presumably use similar approaches across the Neighborhood Keeper program.

- Ahead of time: supporting the evaluation of the data set to determine programmatic blind spots for future participant sensor deployment,
- Real time: frequency of data availability to program participants, data enrichment from necessary external sources, and ability to evaluate detections, as well as contribute indicators for program evaluation.
- After the fact: long-term historical data views that will allow participants to evaluate and analyze data sets across system lifecycles, dynamic regulatory environments, and providing the "look back" capability as indicators are discovered.

The items identified below will assist program participants in consuming the data sets and add value to the overall program.

Parsable Data Field for event.content

The "event.content" field has some lower-level data that could be valuable in reports or visual dashboards. However, in its current form, it is not easily consumed. The field type is a string/text value. To get relevant data out of the field, a post-processing parser is needed to parse out relevant data.

A recommendation would be to make this field a Kibana-supported object field (e.g., a JavaScript Object Notation [JSON]) for parsing inside of the Kibana platform. This would allow participants to performed searches and create reports based on the sub-fields inside of the "event.content" field.

The difficultly of this task will be coming up with a set of shared key:value pairs. However, each "event.content" could be entirely different, so Dragos may need to create a list of keys to store all values contained in this field. They must be careful to keep this list of keys small (i.e., look for shared objects or types of objects between the different "event.content" types).

Attack Group Labels

A data field used to list a threat group or attack group name could not be found. If any of the detections contain behaviors or artifacts from documented adversary groups, it would be useful to add another data field that has a list of groups that have previously used that behavior or artifact. This would allow participants of the Neighborhood Keeper program to use previous attack data to prepare for specific adversary groups and develop indicators of compromise.

YARA Detection Events

INL noted there are "Snort" rule detection events. It would be helpful to also include YARA rule detection events to assist participants in hunting for malicious binaries. This assumes that the Dragos platform does YARA rule creation and detection.

Artificial Intelligence and Machine Learning

Artificial intelligence (AI) and machine learning (ML) can be an essential supporting capabilities for the analysis of critical infrastructure events and potentially the detection of adversarial activities. Appendices A and B provide information on some of INL's AI and ML projects, including recommendations on how Dragos and Neighborhood Keeper could interface with them.

Recommended Modification to Neighborhood Keeper Reports

Dragos provided examples of Neighborhood Keeper program participant reports to INL. The items identified below provide comments and recommendations for the various Neighborhood Keeper reports.

Timeline of Detections

Neighborhood Keeper reports should include a timeline for participants. If there were many detections during a certain period, other participants could check their logs and network captures during the same period to see if any anomalous activity took place.

Data to Blacklist

The data provided in the Kibana instance may not be granular enough; however, it would be useful to provide a blacklist recommendation. This list could include IP addresses, domain names, binary hashes, or names. This type of data should be shareable as it would not reveal anything about the participants' environment, but rather the attacker's or threat group's tactics and methodology.

Applicable Updates or Vulnerability Reports

Although this data is not collected from Neighborhood Keeper participants' environments, it might be useful to include links to recent, publicly released vulnerabilities of OT/IT equipment in the report.

Feedback on the Community Report

The Neighborhood Keeper community report contains two pages. INL recommends including a summary at the beginning that would allow the community report reader to know what they will be looking at as they read the report. Providing clarity on the distinguishing terms for notifications, detections, and indicators would be helpful for participants to avoid any confusion on the significance of each. In addition, it would be helpful to add a general reference of what type of activities are being identified in each of the categories, such as:

- Firmware change
    - Remote firmware updates
    - Local updates
    - Both remote and local updates
    - Only certain reporting devices
- PLC start/stop
    - Local reset
    - Remote reset
    - Local and remote reset
    - Only certain devices or software
- IT/OT traversal
    - IT/OT suspicious traversal
    - All interactive user traversal
    - System to system traversal

As participants consume or leverage the information in reports throughout their organization, they will likely benefit from being able to articulate these details to their teams.

The program overview section provides the number of recipients, the number of notifications processed, and a breakdown by sector of where these notifications are from. It would be helpful to clearly list the timeframe/dates for this.

The detection summary section includes "detections per 12 hours." Including dates would add clarity and be more useful as a historical artifact. As the number of Neighborhood Keeper participants grows, it is possible that the number of categories or detections will become too diverse to be represented in this type of graph. Additionally, the colors in the detection summary graph are similar enough to be difficult to distinguish.

The top 10 detections across the community illustration seems to be a more useful representation of the data. The additional breakdown of detections by severity and then those associated to activity groups is helpful. It is valuable to see some sort of comparison between threat actors and their activities. The bottom ten detections across the community information is also helpful. INL recommends Dragos include

the total number of detections in this section to put these two data points (top 10 and bottom 10) into perspective.

INL found the additional community categories of Firmware Changes, PLC Start/Stops, and IT/OT Traversal helpful; however, they take up a tremendous amount of space in the report. INL recommends these categories presented in a table format. This would also allow Dragos to include other events in the report.

As the Neighborhood Keeper program matures, INL recommends the development of a participant's portal, where each customer would be able to customize or interact with the reports. Participants could then filter elements, modify views, and potentially export the data for further manipulation.

Feedback on the Industry Vertical (Electric) Report

The report provides broad context of what is taking place within the sector; however, it is unclear how the "Subsector Breakdown" is determined as utilities vary in size and function. For example, a utility may only perform distribution functions, or it may encompass generation, transmission, and distribution functions. The report should include both the impact of the function and the function itself. For example, does the Subsector Breakdown consider the total bulk electric system (BES) impact of the function, or is it just the fact that the utility performs that function?

The detection summary includes an Electric Sector Specific Detections high-level categories graph across time. It is recommended that the time frame of the graph be 4 weeks. This allows more accurate observations of the comparison below this graph of last month to current month detections. It is also recommended that the legend not include numbers, as it will create confusion to the viewer who tries to "do the math." For example, the detection summary graph legend provides the following number of detections:

- Electric Generation Detections: 5405
- Electric Transmission and Distribution: 2458
- Electric Support System: 402
- All Detections: 6264

It would seem that arithmetic would be used to achieve the "All Detections" but that does not seem be the case.

The comparison between detections last month and this month is helpful. When possible, INL recommends an annual comparison is included. INL recognizes this graph comparison may not be available until next year.

The report includes the same categories as the community report, but this report also includes the Real-World Attacks category. INL recommends language detailing the approach taken or unique detections classified as an identified firmware change, PLC event, OT traversal, and real-world determination, as it would be useful to the participants. It is also recommended that Dragos include some details of what is in the "electric support system" category (e.g., energy management systems [EMS], SCADA communications systems, OT infrastructure assets, market systems).

Feedback on Vigilance Report

The Vigilance report starts with Your Week by Numbers, then provides weekly event frequency in four categories, including Firmware Change Detections, PLC Commands Detected, Total Boundary Traversals, and Real-World Attacks. The balance of the report then provides detailed information for each of the four categories. Utilities should be able to customize the report to include those categories that are most applicable and appropriate to their environment (e.g., generation, transmission/distribution, oil and gas).

In reference to the Real-World Attack section that is pulled in by the Professional Services and Intelligence Teams, it would be helpful to provide information or links so the utility can gain better insight into any other connected attacks.

As mentioned previously, INL recommends the development of a participant portal with the ability to pull data as needed, tie report notifications internally to a participant so they can search for detections relevant to their organization, and view trends of activity over a larger collection period.

As the program matures, there should be a method to acknowledge or "check out / investigate" a detection while the anonymous entity that the detection maps to is simultaneously working to confirm the activity. In addition, there should be a capability to "close out / resolve" a detection. If it is determined to be a misconfiguration or a false positive, it could be "resolved" as such. If it is determined to be a confirmed event, then the item could be "confirmed" and optionally tied to an information share within the sector Information Sharing and Analysis Center.

Data Sharing Lessons Learned Through INL's Experience With Utilities

Electric utilities subject to the NERC CIP standards have faced some additional considerations specific to information sharing and participation in a variety of programs looking to evaluate OT environment detection. Those concerns often centered around NERC CIP program extensions pertaining to Electronic Access Control or Monitoring Systems (EACMS), granting electronic access, granting authorized unescorted physical access to partners engaged in a program evaluation, and concerns around BES CSI. In many cases it is easier for a program evaluation period to focus on non-CIP facilities, however that limited the ability to evaluate a given program's capabilities throughout some of the most significant sites. That also limited the ability to work through the perceived considerations specific to NERC CIP. The Neighborhood Keeper program has taken an approach that addresses several concerns around access granting and BES CSI.

NERC CIP Compliance

For many utilities that contribute to the BES, complying with the North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection (CIP) regulations are essential. Any solution implemented needs to meet NERC CIP standards. Two regulations create specific barriers to information sharing:

- CIP-004 Personnel and Training: For any data that is declared BES, the utility must be able to control physical access to the electronic storage of that data, which is impossible to do once it leaves their environment.
- CIP-011 Information Protection: For BES Cyber Systems Information (CSI) handling, the utility would need to ensure that the entity they would share information with (e.g., DOE's information protection program) conforms to the requirements within the standard, and the entity would need to provide documentation that the utility could use to demonstrate that compliance to an auditor.

While NERC does not want to stand in the way of security, they have indicated they would need to fully examine the program/environment before considering any waiver or change to the rules. It is important to engage early with the utility's compliance team.

Confidence and Trust Building

Developing trust and willingness to voluntarily share information can be a challenge for entities within a critical infrastructure organization. The following papers provide additional insight on Confidence Building in support of Information Exchange and Cyber Operations.

- **Confidence Building Measures and International Cyber Security**
  - *https://ict4peace.org/wp-content/uploads/2015/04/processbrief_2013_cbm_wt-71.pdf*
- **Conceptual Framework for Cyber Defense Information Sharing within Trust Relationships**
  - *http://gauss.ececs.uc.edu/Courses/c5155/pdf/info_sharing.pdf*
- **Cyber Norms for Civilian Nuclear Power Plants**
  - *https://www.osti.gov/servlets/purl/1358386*
- **Multinational Confidence Building Measures in support of Nuclear Safety and Security**

- https://drive.google.com/file/d/0B2ENqVuD4gMgTXhJSFBNTGREVGM/view?usp=sharing

Recommendations that Could Benefit the Electricity Subsector

Primary areas of recommendation would be:

- Review the recommendations included above in the reporting section
- Provide an interactive participant portal to move beyond static reports
- Include the ability for participants to "check out" a detection and investigate it through the local use of the platform to help "tune" the detections
- Expand the program to many more participants
- Expand the sensors utilized across participant environments
- Provide additional clarity around what is represented in the existing reports
- Expand platform capabilities to allow participant indicator loads
- Provide a "submit for evaluation" feature, to allow Dragos to evaluate indicators for potential inclusion across the program
- As the Neighborhood Keeper program matures, consider additional participant demographics to provide additional filtering capabilities that identify types of facilities impacted, vendors, or protocols being targeted, and potentially sector interdependence issues that could emerge when analyzing a rich data set.

Page intentionally left blank

Appendix A
Use of Artificial Intelligence and Machine Learning

Page intentionally left blank

Appendix A
Use of Artificial Intelligence and Machine Learning

We agree that AI and ML are essential supporting capabilities of any event analysis framework. Our use of these terms conforms to these definitions:

Artificial Intelligence:
In broad terms, AI is a machine's ability to receive data from its environment and subsequently make "smart" decisions using the information. The definition of "smart" regarding these decisions is debated—some scientists believe "smart" refers to thinking and acting like a human while others believe that it means the machine thinks and acts rationally. Regardless, through the process of learning, understanding, and solving, AI provides a way for a computer to reason computationally.

Machine Learning:
Machine Learning is an application of AI. Foundationally, ML is the process of computers using algorithms to learn from data, which represents past experiences. The four primary types of ML are:

1. **Supervised Learning**
   Supervised learning represents Input to Output Mappings. This type of learning requires a large set of data where both A (the input) and B (the correct classification) are available.
2. **Unsupervised Learning**
   Unsupervised learning allows for discovering concepts without identifying the concept classes beforehand.
3. **Transfer Learning**
   Transfer learning learns from one problem and applies this learning to another problem. This is useful in situations where the training set (A, B) has enough samples for training, but the target set does not.
4. **Reinforcement Learning**
   Reinforcement learning is used for Alpha Go and playing video games. It requires large amounts of data so playing video games works well as you can play an infinite number of video games.

Critical Infrastructure Event Analysis using AI and ML:

In general, a successful AI or ML program has two core requirements: a large set of data to analyze and a well-sized set of training data where there are validated matches between the inputs and outputs. Within the ICS event space, researchers should be able to satisfy the first requirement by utilizing the data from all Levels and Zones of the notional ICS architecture shown below (Figure 3). There is an open question of whether to include non-ICS data given that most attacks to date have included an enterprise system bridging into the Operations Management Zone. For this reason, the analysis presented is including these data in the collection set, although this data can be excluded by researchers to narrow their focus.

Data Collection Target Types:

Files: any file contained on a mounted disk or loaded into firmware or memory on any of the devices within the architecture.

Ethernet: all data packets on any of the seven layers of the Open System Interconnection (OSI) model. This would include ICS protocols that are implemented to run across a packet switched data network as well as HTTP and other web protocols that are more typically used for HMI access.

Backplane: all data passing across messaging busses used by ICS control devices to communicate with I/O modules or through a communication module to an HMI.

Serial Bus: all data packets passing across serial bus connections. This would include connections between controllers and field I/O devices as well as USB device connections to any system capable of interfacing with portable serial devices.

Novel Targets

While the focus of this paper is on ICS event analysis and the intersection with AI and ML capabilities, one idea which should be considered is the use of AI and ML to assist in the classification of digital assets.

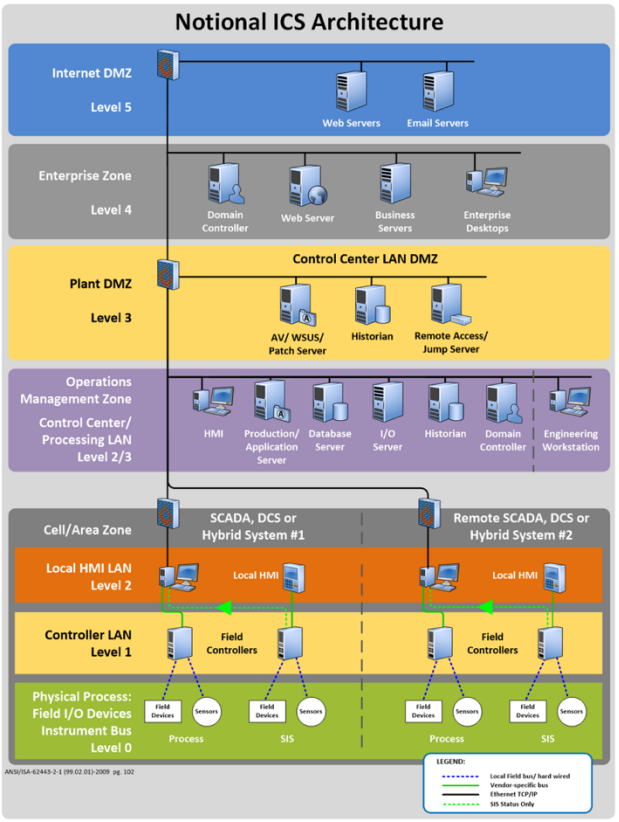

Figure 1: A diagram from ICS-CERT – Recommended Practice: Improving Industrial Control System Cybersecurity with Defense-in-Depth Strategies, September 2016

Critical Digital Asset Assessment:

To discuss cybersecurity within an ICS context, it is essential to understand which assets should be targeted for data acquisition. For example, in Nuclear Power Plants, Critical Digital Assets (CDAs) are components of critical systems (Figure 4), typically perform an safety, security, and emergency preparedness (SSEP) function for a critical system. CDAs could also be assets whose failure as a result of a cyber-attack would result in damage to a critical system.[a] These assets must be protected from cyber-attacks.

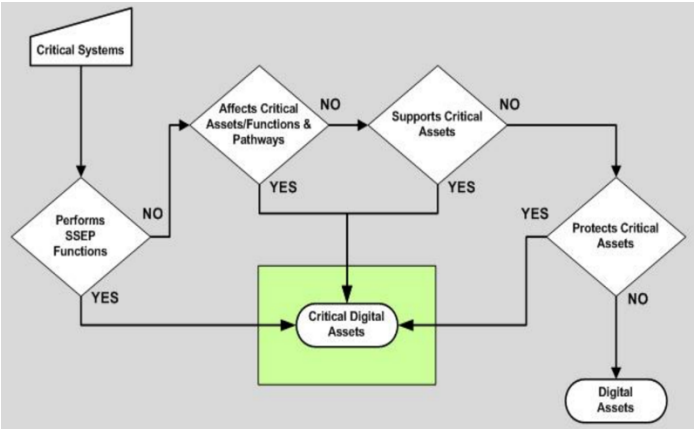

Figure 2: A diagram by the Nuclear Regulatory Commission depicts the evaluation process for determining critical systems.

Data acquisition target generation options

In 2019, researchers from the University of Tulsa proposed a method for identification of acquisition targets within critical infrastructures.

[a] https://scp.nrc.gov/slo/regguide571.pdf

1. Attack Graph Tools: Attack graphs can be constructed to evaluate cybersecurity networks by mapping the route an attacker can take to compromise the system. Using this model, it is possible to identify critical assets and CDAs by expanding on previously known critical assets. To use this technique, there must be at least one asset known to be critical. The output of this algorithm (Figure 5) is a file representing the attack graph containing paths to each state which contains a file with the attributes of the state (asset, qualities). The run time for this method is where *d* is the depth of the attack graph, *f* is the number of fact names (qualities and topologies), and *A* is the total number of assets.

```
exploit expand_critical(a, b)=
  preconditions:
    quality:b,critical=true;
    topology:a->b,physical|
    topology:a<->b,physical|
    topology:a->b,connect_modbus_slave|
    topology:a<->b,connect_modbus_slave|
    topology:a->b,connect_modbus_master|
    topology:a<->b,connect_modbus_master;
  postconditions:
    insert quality:a,critical=true;
```

Figure 3:Output file of Option 1

2. Modified Attack Graph: This method attempts to improve the exponential runtime of the previous method. To do this, the attack graph was modified so the travel through the states models a breadth-first search. This improved the runtime because "all paths through the graph lead to the same final state."[b] The creation of a depth plunge attack graph terminates once it reaches the "bottom" of the graph, so overall it is a better approach to identify the CDAs. The runtime for this method is , where *d* is the depth of the attack graph, *f* is the number of fact names (qualities and topologies), and *A* is the total number of assets.

```
exploit create_cda(a)=
    preconditions:
        quality:a,critical=true;
        quality:a,digital=true;
    postconditions:
        insert quality:a,cda=true;
```

Figure 4: Output file of Option 2

3. Purpose-Built Program: This method does essentially follow the same path as the two above, but it does not generate the attack graph. It can generate a list of CDAs without post processing. The algorithm makes use of two functions: isCritical() and isDigital(), which return if the asset is critical and digital, respectively. Once every asset has been classified as critical and digital, the critical assets are added to a list, then that list is evaluated for all digital assets. The algorithm can be found to the right (Figure 5). The runtime for this algorithm is where *A* is the total number of assets.

```
Let G = {V, E}
V = {V0, V1, ..., Vn}
E = { (vi, vj) when vi, vj is in V}

Function isCritical(v) tests if critical
Function isDigital(v) tests if digital

Critical = []

for v in V:
    if isCritical(v):
        Critical.append(v)

for critical in Critical:
    for (vi, vj) in E:
        if isCritical(vj):
            Critical.append(vi)

CDA = []

for v in Critical:
    if isDigital(v):
        CDA.append(v)
```

Figure 5:Option 3 Algorithm identifying CDAs

The researchers tested the approaches with 54 total assets. They used manual classification as a control. Below is a table with the results. These results show that this is an acceptable way to classify CDAs (Figure 6).

	Manual	Automatic	Adj. Auto
Critical	53	54	53
CDA	43	43	43

Figure 6: Output of test case to classify CDAs

Recommendation:

Given the number of interconnected assets within any critical infrastructure architecture, INL and Dragos could work together with the Neighborhood Keeper customers on identifying data acquisition targets using one of the above methodologies. CDA identification could progress from manual target classification to one that is supported by an automated classification algorithm, such as what is described in Option 3: Purpose-Built Program.

[b] https://ieeexplore.ieee.org/abstract/document/8855281

Page intentionally left blank

Appendix B
National Laboratory Artificial Intelligence and Machine Learning Projects

Page intentionally left blank

Appendix B
National Laboratory Artificial Intelligence and Machine Learning Projects

When assessing which AI and ML techniques and capabilities to integrate into your analysis cell, it is helpful to survey commercial vendors, academia, and National Laboratories. Since this input is being compiled by INL, we will provide an assessment of a subset of programs that may be helpful for Dragos and Neighborhood Keeper to interface with.

Structured Threat Intelligence Graph

Idaho National Laboratory
Link: *https://github.com/idaholab/STIG*

How Does the System Work?
The Structured Threat Intelligence Graph (STIG) provides a threat analysis using Structured Threat Information eXpression (STIX), which is another INL-developed project that provides Unified Cyber Threat Information. Using the information from STIX, STIG provides an easy-to-use visual interface that creates, edits, analyzes, and searches. STIG has a graph database back end to allow highlighting of related threat data objects to focus on the most critical areas concerning the analyst in visual display, rather than reading thousands of lines of code to identify these relationships. STIG is published and free for download on GitHub here: https://github.com/idaholab/STIG.

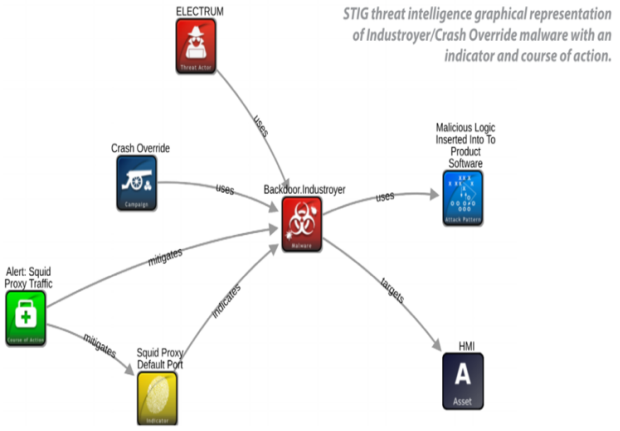

Figure 7: Graphical representation of a malware attack. A detailed description of the symbols used can be found here: https://oasis-open.github.io/cti-documentation/stix/intro.

Connection to AI and ML:
The graph structure of the STIG output provides a stable learning platform for any ML algorithm and can be useful for further analysis and classification of cyber threats. Simply stated, the output of STIG can be used as input for a ML algorithm with a more specific classification goal.

Recommendation:
Using the STIG graph data model, both the scalability and visual discovery of relationships are possible without complex queries. These two qualities allow for easier collaboration between threat analysts and provide the ability to dynamically read and classify the data. In turn, STIG helps to increase the efficiency of detecting and mitigating cyber-attacks on critical infrastructure, which allow defense teams to consistently adjust to the dynamic nature of today's cyberthreat.

Annotated and Translated Disassembled Code

Idaho National Laboratory

How Does the System Work?

The Annotated and Translated Disassembled Code, or @DisCo for short, is an application designed to provide a highly scalable platform that enables automated binary software analysis. @DisCo was created under the Firmware Indicator Translation project, whose main goal was to identify ubiquitous third-party libraries in compiled executables; @DisCo specifically created the ML corpora for this advancement. The application works as follows. First, @DisCo uses the python framework angr to disassemble the binary. After the binary is disassembled and analyzed, its components are stored in a graph database, OrientDB, which creates a complete, searchable control flow graph. In this graph, each block of code is represented by a vector and each function is represented by a group of its block feature sets. Now, by using the graph database, additional information about the binary can be captured and shared.

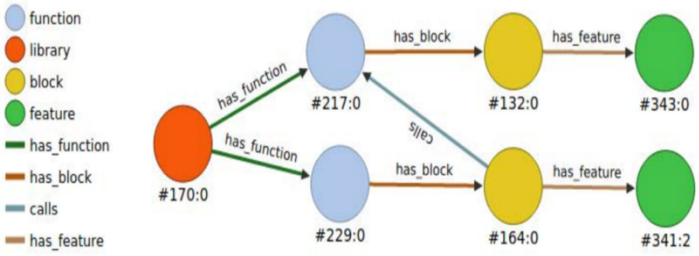

Figure 8: Graphical Representation of a simple "Hello World!" program in @DisCo.

Connection to AI and ML:

@DisCo is designed to provide a collection of ML techniques that is used for its main application of classifying binaries. Essentially, the output of @DisCo provides input for ML and classifier algorithms, allowing it to categorize binaries.

Recommendation:

@DisCo provides a scalable solution for storing binaries in a database and helps identify software components. It can identify vulnerable or altered binaries and system modifications. Most importantly, it augments the reverse engineering process and allows for early detection of issues within the critical system. The Firmware Indicator Translation project used @DisCo's findings to identify unknown function libraries used in certain binaries, which can open doors into new research avenues for scalable binaries and ML.

Autonomic Intelligent Cyber Sensor

Idaho National Laboratory
Link: *http://www.people.vcu.edu/~mmanic/papers/2014/TII14_VolManicLinda_AutoIntellCyberSensor.pdf*

How Does the System Work?
The Automatic Intelligent Cyber Sensor (AICS) is a threat diagnostic system that models the autonomic nervous system in humans—the same system that triggers our "fight-or-flight" response when it perceives a threat. AICS uses ML to detect unwanted or unauthorized entities on a network and allows corporations to avert attacks. AICS creates a model of the system and learns what normal network traffic and patterns look like. Using this information, the system can differentiate between normal behavior and possible threats automatically.

In addition to alerting a defense team of an intruder, AICS can also plant decoys to draw the attention of the attacker away from their intended target. Once the attacker has explored the decoy, AICS "traps" them in the space and allows them to be easily tracked.

Connection to AI and ML:
Instead of depending on operator supplied data, AICS uses a proprietary clustering algorithm (an algorithm that groups data points together based on common information between the two) to learn about the system. Then, AICS continues to learn as it is operating, allowing it to classify new threats that were unfamiliar when the system was first implemented. As a result, AICS can operate with minimal support from human cyber defense workers.

Below, there is a list of steps that represent the researcher's methods to test AICS. Table 1 depicts the anomaly performance rate. The steps below are designed to test the communication functionality of AICS when an attack is carried out on a small campus grid. Each step contains a high-level action followed by the primary functionality to be evaluated. Table 1 shows the results of the anomaly detection for three different streams: 0.3, 0.6, and 0.9, which show the performance with a fixed sensitivity threshold. The IT2 row provides the performance with domain knowledge performing dynamic adjustment of the sensitivity threshold. Overall, the results show that this is a good application of ML.

Table 1: Anomaly Performance Rate

Step 1: Attach sensor to SCG network and initiate system.
- Evaluate NEI host identification model dynamic updates and related IF-MAP messages.

Step 2: Send IP lists to DHP Component
- Test dynamic virtual host creation based on IP list and utilization of NEI host information from step 1.

Step 3: Target Network probes at DHP hosts.
- Verify emulated network presence of devices.

Step 4: Send IP list to IAA anomaly behavior monitoring.
- IAA initiates learning mode, creates normalization and clustering values for fuzzy rule creation.

Step 5: Send monitor message to IAA and initiate network attacks.
- IAA recognizes attacks and sends IF-MAP alerts.

Anomaly Performance			
Threshold	Correct Rate	False Pos.	False Neg.
Stream 1			
0.3	99.8539%	0.1461%	0.0000%
0.6	99.8705%	0.1295%	0.0000%
0.9	99.8788%	0.1212%	0.0000%
IT2 FLS	99.8722%	0.1278%	0.0000%
Stream 2			
0.3	99.9037%	0.1217%	0.0275%
0.6	99.5504%	0.1082%	1.3753%
0.9	99.3799%	0.1082%	2.0079%
IT2 FLS	99.9111%	0.1116%	0.0275%
Stream 3			
0.3	99.8643%	0.2953%	0.0000%
0.6	99.8960%	0.2265%	0.0000%
0.9	99.8960%	0.2265%	0.0000%
IT2 FLS	99.8960%	0.2265%	0.0000%

Recommendation:
The AICS system is specifically marketed towards protecting the energy sector's control systems. In critical infrastructure systems, it is important to be proactive, not reactive. AICS is not meant to replace a cyber defense team, but it instead serve as a tool that defense teams can utilize to aid their classification and deterrence of cyber threats.

Cyber and Physical Anomaly Detection in Smart Grids

Virginia Commonwealth University, University of Idaho, Idaho National Laboratory
***Link**: http://www.people.vcu.edu/~mmanic/papers/2019/RW19_MariWickAmarManic_CyberPhysicalADInSmartGrids.pdf*

How Does the System Work?
This research is focused on the use of a cyber-physical sensor which the authors have named ICS Resilient Security Technology (IREST). The sensor takes as input both cyber events and physical events. Testing was performed on the Idaho SCADA Cybersecurity testbed. The testbed uses hardware-in-the-loop to include industrial-grade hardware and protocols to simulate an ICS environment.

Connection to AI and ML:
The ML models are trained on normal data and use a variation-off-normal approach for determining anomalous activity that the authors classify as disturbances. Supervised as well as unsupervised ML models were implemented and tested within the Anomaly Detection System (ADS). Algorithmic approaches included:

- Signal Temporal Logic Formula to express system properties of their system and support vector machine classifier.
- Unsupervised anomaly detection using Recurrent Neural Networks focused on time-series data.
- Artificial neural network to detect abnormal operation behavior in the system using Dimensional Reduction Techniques.

Test scenarios were divided into three categories: *Normal Operations Scenarios, Abnormal Cyber Scenarios, and Abnormal Physical Scenarios.* This is a common approach given the need for training data necessary to train the classification models.

The IREST sensor contained a packet-sniffer and associated cyber-features extractor for network traffic classification, a DNP3 (Distributed Network Protocol 3) parser that contains the physical device data, and an ML interface for analysis. Their approach was to extract network streams, perform a summary statistic on the window, and combine this with the DNP3 data using a Principal-Component-Analysis method to characterize the physical system behavior.

Decision trees and random forest were trained using a supervised approach to classify normal communication against cyber anomalies. The One Class Support Vector Machines algorithm was trained unsupervised using only normal data. Decision tree and random forest provide comparable results, with 100% prediction score. Their results by algorithm are shown in Table 2:
Table 2: Cyber Anomaly Detection Performance

model	accuracy	precision	recall	f1
OCSVM	0.988	0.987	0.999	0.993
Decision Tree	0.990	1.000	0.863	0.926
Random Forest	0.990	1.000	0.869	0.930

As a reminder, the f1 score is a measure of classification accuracy calculated as .

Recommendation:
This approach should be considered when working through your options for which algorithms and approaches to include in your AI/ML capability suite.

Data-driven Stochastic Anomaly Detection on Smart Grids Using Mixture Poisson Distributions

Virginia Commonwealth University, Idaho National Laboratory
Link: *http://www.people.vcu.edu/~mmanic/papers/2019/IECON19_MarinoWickManic_StochasticADS.pdf*

How Does the System Work?
This research is focused on the characterization and classification of communication within smart grid DCSs. These smart-grid DCSs include interconnections between control systems and traditional IT systems, thus producing the data targeted for collection and analysis for this project. Testing was performed on a simulated SCADA microgrid communicating with real-time automation controllers and a Data Historian using DNP3. Attacks were generated from a dedicated platform at designated times, allowing for validation of ML algorithm observations with event/attack windows.

Connection to AI and ML:
The authors chose not to use an ML black-box approach due to a difficulty in interpreting these models and complicated decision boundaries. What this really means is that an adversary could design their malware payloads to inject data (activity) into the target environment that would defeat the defensive ML models.

The ML models used for this research included data-driven stochastic anomaly detection using Mixture Poisson distributions for packet modeling. The objective is to characterize system behavior by capturing each state (training) and then comparing these normal states with anomalous ones. Learning parameters included a mini-batch Expectation Maximization algorithm due to the large dataset. Algorithmic approaches included:

- One Class Support Vector Machines
- Decision trees
- Random Forests

Data was collected from the testbed and analyzed within the ADS representing system state using directed graphs with nodes corresponding to devices and edges the communications between two or more devices. The adjacency matrices are derived from the collected data for use within the ML algorithms.

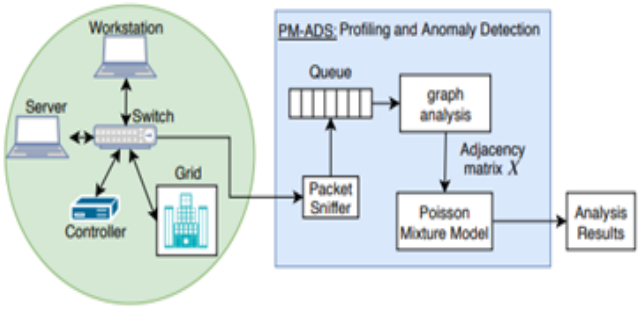

Figure 9: Testbed for Behavior Profiling using Mixture-Poisson Distribution

Normal data was used to train the ADS to characterize and compare normal behavior to identify abnormal behavior. Stochastic Models cumulative density functions (c.d.f.) which is used to test for anomalies by looking at non-covered mixture components. Their results by algorithm were:

Table 3: Anomaly Detection Performance

model	accuracy	precision	recall	f1
PM-ADS	0.991	1.000	0.888	0.941
OCSVM	0.988	0.987	0.999	0.993
Decision Tree	0.990	1.000	0.863	0.926
Random Forest	0.990	1.000	0.869	0.930

As a reminder, the f1 score is a measure of classification accuracy calculated as .

Recommendation:

This approach should be considered when working through your options for which algorithms and approaches to include in your AI/ML capability suite.

Real-Time Cyber-Physical False-Data Attack Detection in Smart Grids Using Neural Networks

Oak Ridge National Laboratory
Link: *https://www.osti.gov/servlets/purl/1426580*

How Does This System Work?
This research explores a ML algorithm to detect data integrity attacks and protect power transmission and distribution systems. The researchers produced a mechanism for cyber-physical attacks detection (CPAD) that infers underlying physical relationships using a multi-sensor approach. Two ML techniques are proposed: neural networks and Support Vector Machines (SVM); however, only the neural networks method will be discussed here as the researchers concluded it was the most efficient and reliable method. The IEEE 30-bus electric power system, which represents a portion of the American Electric Transmission Power System, was used as a test case for the detection method where a connection between nodes (busses) implies a physics-based constraint.

Many of the sensors that researchers are working to protect are already reading data. Because of this, researchers used a collection of those sensor readings to create spoofed data sets. To generate this dataset, the researchers began with a set of bus and branch parameters for the test case. Then, they generated 10,000 new configurations using a probabilistic model generated from the original parameters. For each of these new files, a power flow simulation was run to extract a final state value. Using these state values, a replay attack was run one at a time for each bus, which resulted in 300,000 state vectors in the whole system. Finally, physical features were added for each bus. As a result, this generated a large, diverse dataset of both real and fake data of 10,000 real systems and 300,000 spoofed systems. A neural network-based ML algorithm was then applied to this final dataset.

Connection to AI and ML:
Researchers selected a multi-layered, feed-forward neural networks model in which input nodes present information to the network and output nodes map the decision made by the neural network. For this model, there are 150 input nodes and 31 output nodes (one for each bus and one for “no attack”). Once CPAD has been trained, it helps its operators to identify if, when, and where false-data attacks have occurred.

Following the development of both algorithms (neural networks and SVM), the researchers tested both algorithms on the IEEE 30-bus example. Their results are shown in Table 4:

Table 4: CPAD Test Results

Model Architecture	Accuracy
Single-sensor score	3.2%
No hidden layers (logistic regression)	8%
SVM without physics-based features	45%
SVM with physics-based features	68%
Hidden 20-node layer	97%
Hidden 60-node layer	99%

Recommendation:
This approach should be considered if there is a smart-grid component in data acquisition and there is a requirement to infer physical constraints from raw data while not requiring knowledge of the physical relationships or labeled attack data