Pacific Northwest
NATIONAL LABORATORY

PNNL-33500

Internet of Things Data Characterization Process

Pattern of Life Behavioral Data Study

October 2020

Lisa Campbell
Jereme Haack
Penny McKenzie
Christopher Patrick
Marie Whyatt
Quinn Wright-Mockler
Beverly Johnson

U.S. DEPARTMENT OF ENERGY

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-33500

Internet of Things Data Characterization Process

Pattern of Life Behavioral Data Study

October 2020

Lisa Campbell
Jereme Haack
Penny McKenzie
Christopher Patrick
Marie Whyatt
Quinn Wright-Mockler
Beverly Johnson

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Summary

The HoneyBee™ TARDIS LDRD team completed a data scoping study that identified the initial processes and procedures to baseline the normal and expected behaviors during operability and interoperability of Internet of Things (IoT) device networks. This research is the initial step in developing a process (or methodology) to inform a much broader information framework incorporating machine learning to determine device pattern-of-life which enables the detection of abnormal IoT behaviors on an individual device, as well as in the context of a larger network.

As we look to the growing trends of IoT/IIoT, edge computing, 5G networks, and the ever-changing landscape of IoT/IIoT, it is critical to characterize a device since its interoperability may introduce new vulnerabilities into a network. Characterizing a device will identify its behaviors in use or when it's inactive. The data retrieved from the device need to be analyzed to show information about the device, the connections being made by the devices, how communications are being captured, and what information is being shared with the manufacturer or even third-party entities. Determining what these data mean could provide impactful intelligence of the device, and therefore enterprise, security.

The Internet of Things (IoT)/Industrial Internet of Things (IIoT) world is changing. We are slowly becoming aware of the security implications of IoT/IIoT in government enterprises. These devices are generating up to 80 zettabytes of data which is nearly as much data that has been generated globally up to this point in time. Most of that data is not being handled in a confidential way, and organizations are not able to identify the device leading to a lack of data confidentiality. The lack of confidentiality means that proprietary, patient health, and personal user information and organizational networks themselves are at risk. The implications of third-party access to this information/networks further expands the vulnerability. This risk is compounded by the fact that there are no common or proper authentications for IoT/IIoT. Therefore, when deployed, these devices introduce new and novel vulnerabilities of which an enterprise may not be aware. Most of the data being generated by these devices are also unencrypted and the data is not being handled in a secure manner. The confidentiality of the data is missed because manufacturers find it very difficult to enforce encryption on devices that do not have the proper computing power to support encryption. This lack of protection in turn compromises the integrity of the devices themselves. How do you know that the data being sent to the device of the manufacture is coming from a trusted source? The integrity of the device is extremely critical. For example, as users make decisions related to the data and personnel information being sent about patient care and what treatments to use or the direction that an autonomous vehicle needs to take to get to its destination, it is easy to recognize that ensuring the values of the associated data have integrity is critical to human lives.

Upon formulating the data points that needed to be captured in this study, we have gathered some interesting results. Most information being shared can be traced back to a cloud instance. In the case of some IoT/IIoT devices that may have multiple applications associated with it, we found that just one source of data information being shared and gathered was not the case. The results formulated by the team addressed how the devices were communicating, what the network traffic looked like, and any additional connections that were being made with other lab equipment. What we found was that the devices chosen had unique properties individual to that specific device and that the data being sent and received by the device was not exactly for what it was designed.

This report includes the data communications that were captured from two devices by the network monitoring and analysis tools used by the team. We also gathered open-source information to correlate

information that was advertised and documented with the user manuals with each device. Upon further investigation into the 3rd-party communications, it was determined that traceability intelligence might have significant value to the device owner. We believe that advancing this study through additional IoT device data captures to build pattern-of-life behaviors, device identification and classification and extending it to developing machine learning techniques to capture and analyze data will enhance the intelligence analysis of identifying and securing devices. The results from the long-term research will provide additional insight for enterprises to further protect themselves and their networks.

Acronyms and Abbreviations

Amazon Web Services	AWS
Comma-Separated Values	CSV
Common Vulnerability Exposures	CVE
Domain Naming Service	DNS
Elastic search, Logstash, Kibanna	ELK
Energy and Environment Directorate	EED
Federal Communications Commission	FCC
Industrial Internet of Things	IIoT
Information Technology	IT
Institute of Electrical and Electronics Engineers	IEEE
Internet of Things	IoT
Internet Protocol	IP
Media Access Control	MAC
Network Time Protocol	NTP
Open Systems Interconnection model	OSI
Organizationally Unique Identifier	OUI
Radio Frequency	RF
software defined radio	SDR
Television	TV
Transmission Control Protocol	TCP
User Datagram Protocol	UDP
wireless intrusion detection	WIDS

Contents

Figures

Tables

1.0 Introduction

The HoneyBee™ TARDIS LDRD team completed a data scoping study that identified the initial processes and procedures to baseline the normal and expected behaviors during operability and interoperability of Internet of Things (IoT) device networks. This research is the initial step in developing a process to inform a much broader information framework to enable the detection of abnormal IoT behaviors on an individual device, as well as in the context of a larger network. Exercising this method to identify impacts of network interactions using defined patterns of behavior can improve the ability to protect intricate systems, both internal and external to the networks, from cyber attacks. Hence, it has the potential to alleviate cyber security risks. As the data are collected and analyzed, the intelligence may be leveraged to identify unique and unexpected conclusions.

Data capture and device defining activities will inform the characterization of IoT device Pattern-of-Life (POL) events which are comprised of device identification, communications, and device interactions. This process helps define some of the criteria necessary for device POL fingerprinting (e.g., data capture, identification of physical attributes), as well as determine the initial classifiers that are foundational to the identification and development of POL characterization.

Currently, this capability, which is critical in order to detect anomalous devices and behaviors on specified networks, does not exist mainly because of the dispersed number of devices and manufacturers. However, it is recognized as an emerging need because of the security impacts device interoperability affects the public, private, industry and government.

This study focused on a limited number of devices and a limited assessment. The team concentrated on fingerprinting two devices: the Yale Keyless Lock with Bluetooth and wireless capabilities and the TLC 4K Smart Television with wireless and ethernet capabilities. The assessment process mainly ***focused on specific data captures*** and did not include physical, cyber-physical, nor the step-by-step evaluation of a device. However, the team did capture a secondary component of what that process may entail. The results from this study helped the team answer the following questions to identify POL:

- Does the device send data back to the manufacturer/provider and what data is being sent?
- What function does application/device perform?
- What does the unique identifier look like and what is the purpose of the identifier?
- How do can you classify this identifier for identification of the IoT device?
- What are key identifiers that fulfill like uniqueness (i.e. scalability, privacy, etc.)?
- Are the data from device interactions providing additional opportunities for intelligence gathering?

To address the security risks of having unknown devices or atypical device behaviors present on a network, it is important to establish POL properties from physical and data characterization of devices. These POL events determine the behavioral model that is comprised of device identification, communications, and device interactions. The behavior model is a baseline component for developing the information framework that will dictate the mitigation strategies for organizations that implement the tool. This research will enable the detection of abnormal IoT behaviors on an individual device, as well as in the context of a larger network.

1.1 Current IoT Security Challenges

Significant effort has been devoted to protecting high value and critical national security assets from physical threats, cyber threats, and operational threats. However, a new threat is quickly emerging that has not yet garnered a lot of attention: the Internet of Things (IoT). The interconnected nature of a multitude of devices does not necessarily need the internet to communicate just an active connection to a structured or ad-hoc network. In 2014-2017 mobile phones contributed to most of the wireless communication market and communicated primarily over a cellular network. In 2018 the number of IoT devices surpassed the mobile phone in sheer numbers and implemented a wide range of communication modes (Wi-Fi, Bluetooth, Infrared, Radio Frequency (RF), etc.). In 2025, it's projected that almost 76 billion devices will be interconnected and actively communicating over a wide variety of proprietary and open communication protocols.[1] These estimates are very conservative. The rollout of the new 5G networks and the growth of connected devices may double or triple the number of interconnected devices from previous estimations. This leads to a global population of devices that are designed, used, and maintained with vastly different levels of importance placed on cyber, physical, and operational security requirements.

One challenge that many organizations face is the evolving nature of their networks because of interconnection of electronic devices. As more of these devices become network-enabled-and more organizations add those devices to their internal and external networks (either intentionally or unintentionally), the traditional information technology (IT) network architecture changes. The network changes can create new attack pathways that security professionals are unaware of until a compromise has occurred. Another challenge is that most devices are not easily recognizable when connected to these networks. This unfamiliarity can create security issues that organizations are not ready or able to mitigate in a timely manner.

1.2 Approach

To address device security on IoT device operability and interoperability (i.e., device connection and communications), it is necessary to formulate and apply an assessment methodology that will provide behavioral analyses of IoT devices through data and information gathering. Elements that are critical in mitigating IoT device vulnerability are:

- Knowing device security – or no evidence of being unsecured
- Connected or reaching out (in today's environment, it should be a normal assumption that a device is reaching out)
- Identification of devices and how they are used to reach back to you
- Analyzing who is gathering the data
    - Why?
    - Who?
    - What?
    - How?

[1] Published by Statista Research Department, Nov 14, 2019, Retrieved January 20, 2020, https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

    - What does it provide?
    - How are you vulnerable?

Our approach to support the above listed elements was to identify the initial process IoT device data characterization. These elements are part of the foundational process, and additional data elements are required to create a comprehensive POL and IoT device identification. We maintained a limited scope for this Tardis study. This process focused on:

- Physical identification and assets
    - Manufacturer information
    - MAC address
    - Connection
    - Listed communication
    - Open source recon
- data capture and analysis
    - device identification
    - device categorization

As the team moved forward with defining the initial steps in the process, they referenced other assessment methodologies such as the CORRAL Assessment Methodology. However, full encapsulation of additional processes into the IoT evaluation process was beyond the scope of this effort. Therefore, the team focused on determining if each device has a unique set of properties or data points that can classify the device based on network interaction, human interactions, and stand-by mode. The team chose two IoT devices that have multiple network communication capabilities (wireless and Bluetooth) to identify if the communications would influence the identifier for the device.

1.3 Devices

The two devices deployed for this study—a smart television and the keyless lock— were chosen as examples of devices which utilize diverse communications and are installed in numerous different environments. These devices provide examples of the many different manufacturers that each device type can have.

The television employed—TCL - 55" Class 4 Series LED 4K UHD Smart Roku TV Model # 55S425—works with Amazon Alexa, Google Assistant, built-in dual-band Wi-Fi, and an Ethernet port.[1] The locks employed— two Yale Keyless Locks, Serial # H6NJ9000093 and H6NJ9000120—have the following configuration capability or restrictions as stated by the manufacturer:

- Yale smart locks, along with SmartThings Hub, allow the user to remotely lock and unlock their door form anywhere using a downloaded application from the manufacturer website.

[1] https://www.tcl.com/us/en/products/home-theater/4-series/tcl-55-class-4-series-4k-uhd-led-roku-smart-tv-55s425, 10/14/2020

- Compatible Technologies Z wave (800-900 MHz radio frequency range) Bluetooth
  - Amazon Alexa
  - Samsung smart things
  - Wink
- Included is the information for programming a master code on the lock. This can be between 4 to 8 numbers, respectively. It can have up to 30 users of the lock. the lock can be reprogrammed by taking out the battery and holding down specific buttons to factory reset the device. The warning states that the lock may be defeated by forcible or technical means.
- This device is licensed as RF through the Federal Communications Commission (FCC). These devices may cause interference with radio services and television signals.
- CAUTION: Prevent unauthorized entry. Since anyone with access to the back panel can change the user codes, you must restrict access to the back panel and routinely check the user codes to ensure they have not been altered without your knowledge. The use of a master code can help protect your system's settings.
- WARNING: This Manufacturer advises that no lock can provide complete security by itself. This lock may be defeated by forcible or technical means or evaded by entry elsewhere on the property. No lock can substitute for caution, awareness of your environment, and common sense. Builder's hardware is available in multiple performance grades to suit the application. In order to enhance security and reduce risk, you should consult a qualified locksmith or other security professional.
- Communication Protocols used are Zigbee, ZWave.

1.4 Network Architecture

It is important to understand an IoT network architecture in order to consider network configuration, data communications, and device interactions. The complexity of an IoT network can be daunting because of the magnitude of devices found on a basic network. The conceptual design of the Internet of Things Common Operating Environment (IoTCOE) provides this realistic understanding of an IoT infrastructure. Even within the complex environment, the IoTCOE team mapped the basic architecture design of the IoT ecosystem in a controlled lab environment. The IoTCOE is made up of numerous different elements that include sensors, protocols, actuators, cloud services, and other layers. The team stood up each IoT component to distinguish and track the consistency of the baseline IoT network. Thus, this configuration enables reproduction and secure network deployment of newly acquired devices to support a dynamic and scalable lab environment. Figure 1 illustrates that the basic network architecture consists of the IoT devices layer (sensors and actuators), the IoT gateway layer and acquisition systems, and the IoT platform layer (edge IT and cloud). Fundamentally, the architecture includes functionality, scalability, availability, and maintainability. Figure 2 illustrates the communication pathway from the device to external database or cloud infrastructure that is part most commercial IoT devices.

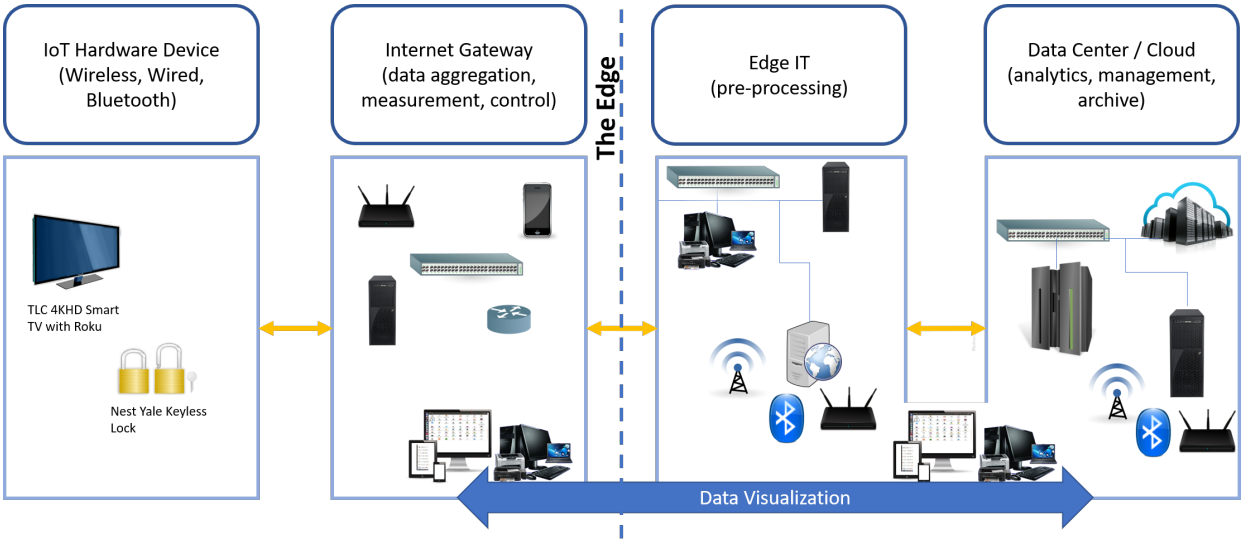


Figure 1 IoTCOE Network Diagram

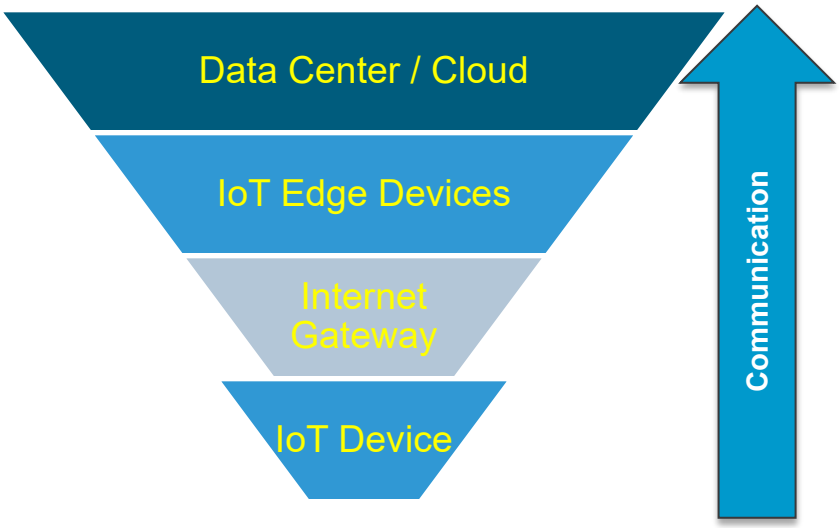


Figure 2 Network Communication Layers

2.0 Analysis Process

When identifying the process to perform device Pattern of Life (POL), it is important to know all aspects of the devices including physical attributes. For this study, we captured a subset of what would normally be captured from a device. As the devices were being unpackaged for deployment, the team noted manufacturer information, serial numbers, FCC identifiers, and any other unique identifiers assigned by the manufacturer within the documentation that could be useful during the analysis. Any additional serial numbers, FCC identifiers, IP addresses, media access control (MAC) addresses, or other unique identifiers that were located on the individual hardware pieces were also noted.

In order to perform the analyses of the data captures from the two identified devices in the IoTCOE device data study, the environment and devices needed to be prepared and deployed correctly. Devices requiring a wired connection are connected to a Cisco switch that has access to a designated network. The wireless devices are connected to the same designated network with a network key provided by network services at PNNL.

The placement and use of IoT devices change network baselines (and sometimes very rapidly). When deploying a large number of devices (designed to be easy to install), keeping up with an ever-changing baseline can be extremely difficult. Additionally, there is no current capability that identifies what is normal, expected behavior for a device. Different sources and manufacturers use different protocols and languages and there is no common structure or format. Aggregating the data would not necessarily be difficult but normalizing it can be.

To capture the network data, multiple COTS/open-source tools were used. PNNL's Splunk instance constantly captures data and needs no special permissions or activation to do so. The other tools, Wireshark[1], Kismet[2], TCPDump[3], and Colasoft Capsa Enterprise[4] needed to be deployed and configured to monitor network traffic by a team member. These actions were performed at various times of the day and in various ways to simulate real-life use as much as possible on the chosen devices.

2.1 Assumptions

This study limited scope to a subset of the physical attributes and the data communications of the two identified devices.

[1] https://www.wireshark.org/
[2] https://www.kismetwireless.net/
[3] https://www.tcpdump.org/
[4] https://www.colasoft.com/

3.0 Data Collection and Characterization Process

As identified in Section 1.0, this study concentrated on data captures. The purpose of the data collection and characterization process is to identify unique properties that include connectivity, network communications, and manufacturer information in commercial-off-the-shelf interconnected devices, specifically the TCL 4K UHD Smart Roku Television (TV) and the Nest Yale Keyless Door Lock. The initial process includes a comprehensive set of procedures for assessing the network communications, data exchange, and use behaviors of these deployed devices in the IoTCOE. The testing team used open-source and commercial data capturing tools along with research of vulnerabilities published in the National Vulnerability Database. The results of the data collection and characterization process are recorded in worksheets (Appendix A) along with information that identifies vulnerabilities and unique identifier information about the IoT device. Included in the report are network configurations and security procedures used for safe deployment of the device within IoTCOE.

For the purposes of the TARDIS LDRD Data Characterization effort, the project team employed a select set of the process procedures to keep the scope controllable. The modified evaluation steps encompassed both the data and device identifier collection. The evaluation procedures are listed in Appendix A.

Data gathering activities included for the Tardis, but are not limited to, the following:

- Full inventory list of devices deployed
- Open-source information gathering on Mac addresses, Passwords, Admin credentials, and other security options for devices deployed
- Deployment of Splunk instance on network for capture communications to outside sources
- Deployment of Wireshark to capture internal network communications
- Deployment of Kali for the use of Kismet to capture Bluetooth communications.

Device assessments would use one or more of the following assessment methods:

- Review, analyze, inspect, or observe one or more assessment artifacts[1] that are attached to the devices.
- Perform technical tests, including functionality testing and network captures on device components using automated and manual methods.

Features that would be tested

- List the features of the software/product to be tested
- Provide references to the requirements and/or design specifications of the features to be tested

Features would not be tested

- List the features of the software/product which will not be tested
- Specify the reasons these features will not be tested

Test environment

- Specify the properties of test environment: hardware, software, network, etc.
- List any testing or related tools used during device identification

Risks

- List the risks that have been identified
- Specify the mitigation plan and the contingency plan for each risk

[1] Artifacts are items that get left behind based upon the activities of the end user of the device.

Assumptions and dependencies

- List the assumptions that have been made during the preparation of this plan
- List the dependencies

3.1 Data Collection Process

In order to effectively test the chosen devices, the team created a repeatable process that captured any preliminary data that could give insight into some of the challenges (e.g., compatibility issues) presented by each device. For instance, the Yale keyless lock required a smart hub to communicate with an app on the Samsung S10 smart phone in the lab which required the team to deploy the Google Home Smart hub for communication. After determining normal operations, the steps the team used for the initial process data collection process are as follows:

1. In what kind of environment will the device be operating?
    a. Other devices it may interact with
    b. Enabled/disabled protocols
    c. Known security settings that may or may not be used
    d. Other device specific items
2. Ensure that testbed can support the device
    a. Network access
        i. Wired or wireless?
    b. No known conflicts with other devices in testbed
3. Literature search
    a. Device
        i. Version differences
        ii. Change logs
    b. Associated devices (devices that may talk to/receive communications from it)
    c. Third party libraries that may be compiled within device
        i. Identify code
4. Device testing scope
    a. What activities/communications are considered inbounds?
5. Device integrity check
    a. Verify authenticity of device
        i. Compare hardware/software to known normal configurations
6. Baselining to understand normal
    a. What does normal look like for an average user?
    b. What do normal communications look like?
7. Develop Test Plan

3.2 Data capture

The data captures are essential in the extraction of information of data exchanges between the IoT devices and other peripheral devices. The team collected relevant information about the network interactions that included third-party connection information that was broadcasting network data. The software tools listed below were utilized during the data assessments and the various information that the team was able to collect during the data study. Using multiple tools allowed the team to gather

multiple data points as well as cross referencing traffic and all screen captures are featured in Appendix C.

The initial software toolset for data captures included:

1. Kismet was used to capture Bluetooth communications from the two identified devices.
    a. Kismet is a wireless network and device detector, sniffer, wardriving tool, and wireless intrusion detection (WIDS)framework. [1]
2. TCPDump was used to capture all wired and wireless network communications.
    a. TCPDump is a network packet analyzer. Adding various options to native commands allows users to search for packets with specific characteristics or that match Boolean search terms.[2]
3. Wireshark is constantly used in IoTCOE to capture network traffic on an ongoing basis creating a true network communication baseline.
    a. Wireshark is a network protocol analyzer. It allows users to view the contents of network packets in a graphical format. It also allows for captures from different tools to be imported, provided it is in a format that Wireshark can decipher.[3]
4. Colasoft Capsa was used for the added functionality of network communications breakdown and ease of use.
    a. Colasoft Capsa is a portable network analyzer application for both LANs and WLANs which performs real-time packet capturing capability, 24x7 network monitoring, advanced protocol analysis, in-depth packet decoding, and automatic expert diagnosis. [4]
5. Splunk was used to break down the structure of the network communications by using customizable correlations, searches and visualizations of all the network data captures during the data study.
    a. Splunk is a platform that allows the investigation, monitoring, analysis and act with any structure, any source, any timescale, any insight, and any action.[5]

Expanded studies would include power consumption data in addition to network traffic. While Splunk provides much of the network data, exporting that data to a comma-separated values (CSV) and importing it, in addition to any power data, into an Elastic search, Logstash, Kibana (ELK) stack can provide a single data aggregator better data manipulation. A dedicated ELK stack would also allow for additional network data to be imported once missing data points were identified. This would help provide a more complete picture of a device's network traffic.

Using each of the identified tools, the team analyzed the data captured to understand what data points are relevant to the identification of the devices tested. This included:

- Dominant Protocol Analysis
- Open Systems Interconnection model (OSI) Protocol Layers – dominant protocols that can be used to extract features – port intervals, packets size, packet quantity, availability time, inter-arrival time, cumulative count (application layer, Transport layer, network layer, data link layer)

[1] https://www.kismetwireless.net/ 10/14/2020
[2] https://www.tcpdump.org/manpages/tcpdump.1.html
[3] https://www.wireshark.org/
[4] https://www.colasoft.com/capsa/
[5] https://www.splunk.com/en_us/about-us/why-splunk.html

- Category device wireless/wired
- System ports registered ports dynamic ports Transmission Control Protocol (TCP) port 443 is used by all devices
- User Datagram Protocol (UDP) packets
- Domain Naming Service (DNS) and Network Time Protocol (NTP)

4.0 Results

The initial experimentation shows promising results to further the development of the data characterization process, pattern-of-life device detection, and device identification using the analysis framework. Understanding the interplay between the device manufacturer, third-party connections, and network interactions can lead to new techniques for detecting and classifying IoT devices to support enhanced cybersecurity processes for a network.

We sought to answer the following questions for each device:

1. How does the device communicate with the manufacturer/provider?
    a. Unknown connections that can lead to unknown network connection and compromise from unknown sources.
2. What is the main function of the application/device?
    a. Difference in the functions can show possible malware, abnormal behaviors, or compromise.
3. What does the identifier for the device look like?
    a. Each device has a unique set of properties that can categorize the device and identify otherwise unknown devices connected to a secured environment.
4. How do you classify the identifier for the device?
    a. It is important to be able to create linkages between a manufacture database, cloud provider, vendor cloud instance, to enable analysis of the devices, and comparisons between devices. These affiliations can identify ultimate parent devices, subsidiaries and generally how device and manufacturers relate to each other.
5. How does the identifier fulfill the uniqueness of the device?
    a. Any IoT device that is intended to affect the structure or function of the network could contain a distinct piece of code that can be used.
6. What other intelligence did we gather?
    a. Investigation of unexpected third-party communications and their impacts.

The team was able to identify some unique properties about each device from the network traffic captured. The Mac address for each device can be traced back to third-party web interfaces.[1, 2] Researching the Mac address on the internet resulted in identifying that the device domain names are under control of the manufacturer not the individual registrant leaving the device networks open to potential cyber activities. Since each device has different communications and additional capabilities, the team initiated the initial search for these types of connections that are not easily understandable by users. This information also showed supply chain issues that could arise from MAC address correlation with Institute of Electrical and Electronics Engineers (IEEE) Standards Organizationally Unique Identifier (OUI) list and third-party connections with the Public Suffix list.

The team was not working on campus during the pandemic shutdown; therefore, we were able to gather real baseline interactions of the devices in an idle state. This is important as many of the IoT devices have an "always listening" capability and the network interaction in this listening state can skew the baselining results. The *"idle state"* baseline that the team compared resulted in being able to identify changes of the network interaction and devices behaviors when in use. The device *"use state"*

[1] http://standards-oui.ieee.org/oui.txt
[2] https://publicsuffix.org/list/

added to the complexity of the network interactions, but the "*idle state"* baseline was able to identify the third-party network connections that the team was not aware of before the study was started.

Roku Television (TV)

The Roku TV is designed to provide access to various video and audio streaming services. In addition to expected communication with Amazon Web Services (AWS) and Roku.com, traffic to applications that were installed on the TV but not in use were seen sending and receiving third-party information and keep-alive packets throughout the capture period. The identifiers used were IP address and MAC address. Both addresses serve to distinguish this device from others in network captures that collect data from multiple devices.

Additional intelligence may be gathered about device communications through more in-depth investigation. The team performed a brief evaluation of the communication of the inactive applications and found that it led to a discrete 3rd-party source. We believe that further investigations for this type of device interaction may bring more cyber intelligence about the device network. These next steps, although limited in this study, in fingerprinting device behaviors are some of the value that extend POL knowledge and upon which action may be taken.

Yale Keyless Lock

The Yale Keyless lock is designed to provide a remote access capability to a traditional door lock. Traffic from the door lock is routed through the Smart Things hub that is connected to the network via Ethernet. The identifiers used were IP address and MAC address. Both serve to distinguish this device from others in network captures that collect data from multiple devices.

When the team started the study there was an assumption that the communications and connections could be easily identified. It was found that this was not the case. The TCL Smart TV was broadcasting information to the roku.com website and third-party connections that included Spotify and Amazon Web services. The team also found that a network communication connection was intentionally being made to an unknown address that had an active, open communication line to the TV and the network.

There was also an assumption that the device would need to be configured with a new username and password during deployment. As we found that with the Smart TV and the Yale Keyless lock, you are required to supply username and password information. What was interesting is that there was no standard for this type of configuration. This means that a user could easily put admin/admin or me/1234 as their username and passwords for their accounts. There is no accountability for the user or the manufacturer for basic cyber security controls.

5.0 Conclusion and Next Steps

In Section 4.0, we defined the findings from the limited scope device data characterization. It showed the value in characterizing a device, even within the limited scope, to determine its POL by addressing specific questions of interest. The steps taken are an initial effort to formulate a process that could lead to an information framework to address cyber risks associated with IoT devices. It is clear that a greater effort is needed for completion of even the two devices we characterized. Although we know that additional steps need to be taken within the process, there are steps that may not yet be identified in characterizing an IoT device.

Of significant value to expanding this POL process is the ability to detect devices; create new datasets (from both current and new devices); identify future threat signatures; determine network level cyber threat detection, prevention, assessment and response; and multi-layer situational awareness through ML/AI networking and cyber aspects. Figure 3 is representative of the processes that needs to be captured for a more complete understanding of device POL.

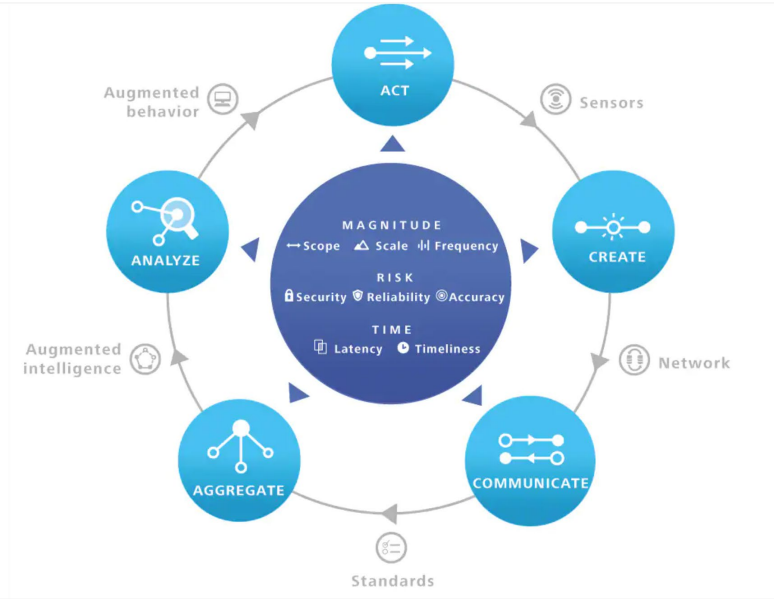


Figure 3 IoT Security Lifecycle

Although the next step options are broken out into a phased approach, it may require that various components (or parts of the components) in both phases (e.g., intelligence gathering) in order for the process to be more effective.

5.1 Phase 1 Development

Discover, assemble, curate, wrangle, or generate data sets at any classification level that will directly enable R&D to develop data-driven methods to identify IoT devices by network communications and other means. The team will continue to identify rogue devices and provide a POL behavior of deployed IoT devices to better identify them and their locations. This work will provide insights into possible cyber-attack vectors that an organization may not have previously been aware. In addition, this research paves the way for creating visualizations of curated data for POL analysis.

1. (Enhance) **Assessment Evaluation:** Evaluate other IT and OT assessment methodologies both in-house developed and external to select key components for incorporation.
    a. Coordinate/integrate IT/OT assessments and network tools
    b. Identify other tools needed to fingerprint and analyze a device
    c. Developed by PNNL or COTS
2. (Physical) **IoT Device Evaluation:** Identify physical components, attributes, and manufacturer specifications of a device, including chipsets, firmware, etc.
3. (Detect) **IoT Device Detection:** Leverage additional IoT device detection technologies to establish an IoT device baseline in a designated area.
    a. Scale to more devices
    b. Incorporate emerging technologies
4. (Store Data) **Create Additional IoT Datasets:** Gather detected IoT device information and activity to create additional IoT datasets over space and time (real-time experimental dataset examples: 1 hour, 2 hours, 4 hours, 12 hours, 24-72 hr. experiments).
    a. Power consumption data
    b. Additional network traffic
    c. Static vice dynamic behaviors (changing environments)
    d. Integrate with physical security, homes, behavioral -based ML on health monitoring and first response, etc.
    e. Data control capability incorporating VOLTTRON™ platform for collecting and storing data
5. (Data Processing) **IoT Data Aggregation:** Further develop a pipeline and IoT common enumeration language for aggregating and collating disparate IoT data sources into common data format.
    a. Additional and automated testing
    b. Automate processes to include ML (e.g., CHISSL)
        i. Identify data and devices identifications algorithm for unique identifiers
        ii. Develop ML algorithm for normal and abnormal device behavior
6. (Data Viewer) **IoT Common Operating Platform (IoT COP) Viewer:** Visualize and present IoT device data to users (e.g. non-cyber analysts / security specialists) in a geospatial context.

5.2 Phase 2 Development

Develop and describe new and untapped IoT POL indicators of propagation activity. This phase 2 effort will include identification of the data source and discussion of the computational analysis method to extract the signature and a potential, realistic concept for implementation of the method as an operational capability.

1. (Sensemaking) **Identify IoT Device Threat Signature:** Continue to identify multiple device interactions, trends, digital footprints, and pattern-of-life characterizations of IoT computing environments.
    a. CVEs
        i. Verify
        ii. What can the identified vulnerability mean to the network
    b. Intelligence gathering
        i. Internet searches about the devices before and after purchase and deployment
        ii. Investigate third-party connections (known and unknown) for active and inactive devices

2. (Modeling) **Experimentation and Use Case Development:** Discover and identify previously unknown and untapped IoT POL signatures for the purposes of enhancing threat activity characterizations:
3. *Use Case 1* – Evaluate and Test additional IoT device sensor data: Compare the data to existing geospatial datasets.
4. *Use Case 2* –IoT Device Activity Characterization: Develop cataloged, machine-readable (i.e., high-Technology Readiness Level [TRL]) datasets and lists of IoT identified sources/data streams and a discussion of their relevance.
5. (Workflow Integration) **Research to Operations Requirement Scoping Study Report:** Evaluate developed prototypes and identify requirements for compatible transitioning into:
   1) Software/Tools that exist in the development or production environment and 2) Standalone prototype outputs that can integrate into analysis workflows. The intent of this outcome is to establish a common foundation of shared components across key systems, which will make them interoperable "out of the box" rather than today's model of leaving the integration for last.

5.3 Data Collection with VOLTTRON™

The PNNL-developed VOLTTRON™ platform provides an extensible solution for collecting and storing data from supported IoT devices (such as the Ecobee thermostat) along with more traditional building sources such as BACnet based building management systems or Modbus devices. VOLTTRON™ also provides an environment for executing energy efficiency applications being developed by Energy and Environment Directorate (EED) researchers including the Connected Homes project which is looking at balancing residential energy savings with homeowner comfort. The extensible driver framework in VOLTTRON™ provides a potential method for adding new devices to participate in data collection and application execution. This capability also extends options for investigating the security implications of integrating devices and applications in a potentially insecure network environment.

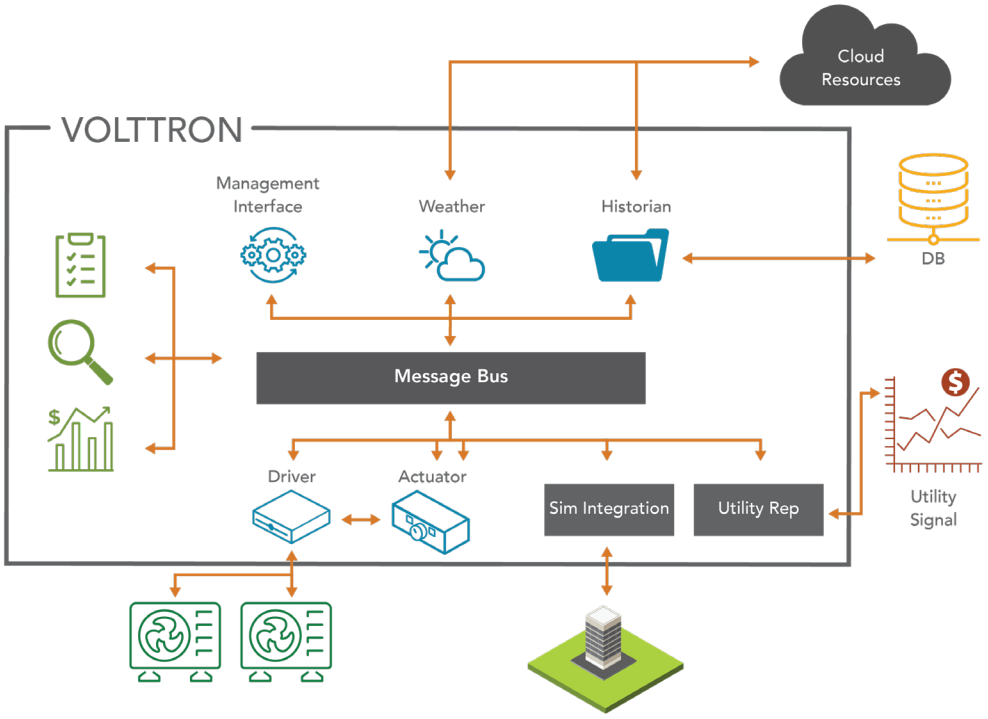


Figure 4 Overview of the VOLTTRON platform. Drivers interact with devices making data available to applications for analysis and historians for storage.

5.4 Analysis with CHISSL – Human Assisted Machine Learning

One tool for analyzing the data is the CHISSL human assisted machine learning tool developed by researchers in the Visual Analytics group. CHISSL can group collected data by similarity and then present these groups to the analyst for verification or allow them to make changes which the system then uses to update its groups. Initially developed for high level user actions such as log in, application usage, network activity, etc., CHISSL has been applied to building data collected by the VOLTTRON PNNL Campus deployment. Researchers used CHISSL to define groups of sensor data streams which showed typical behavior patterns and highlight anomalies. These techniques can be similarly applied to the data being collected by HoneyBee™ to define groups of behavior patterns for devices and identify off-normal activity.

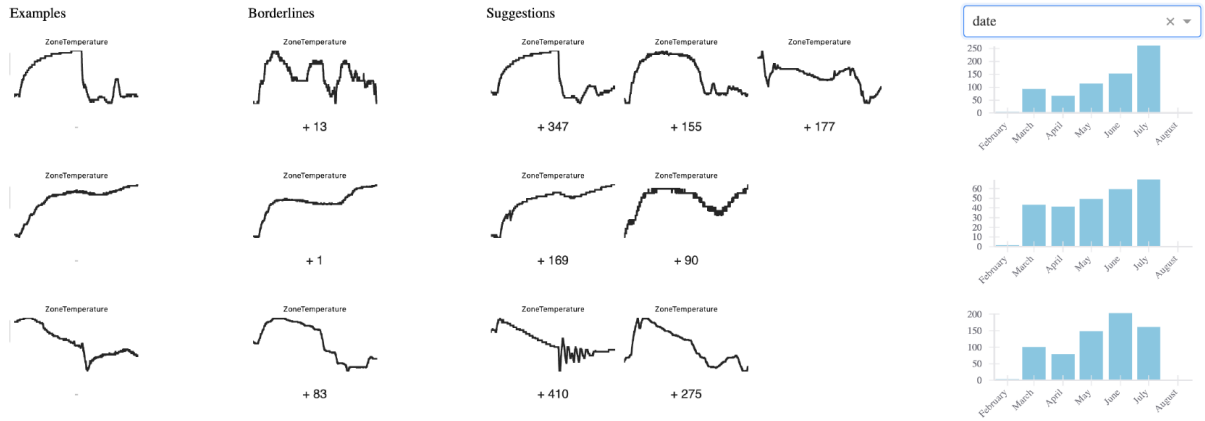


Figure 5 CHISSL Interface allowing the user to provide feedback to the system for suggested groupings of time series temperature data

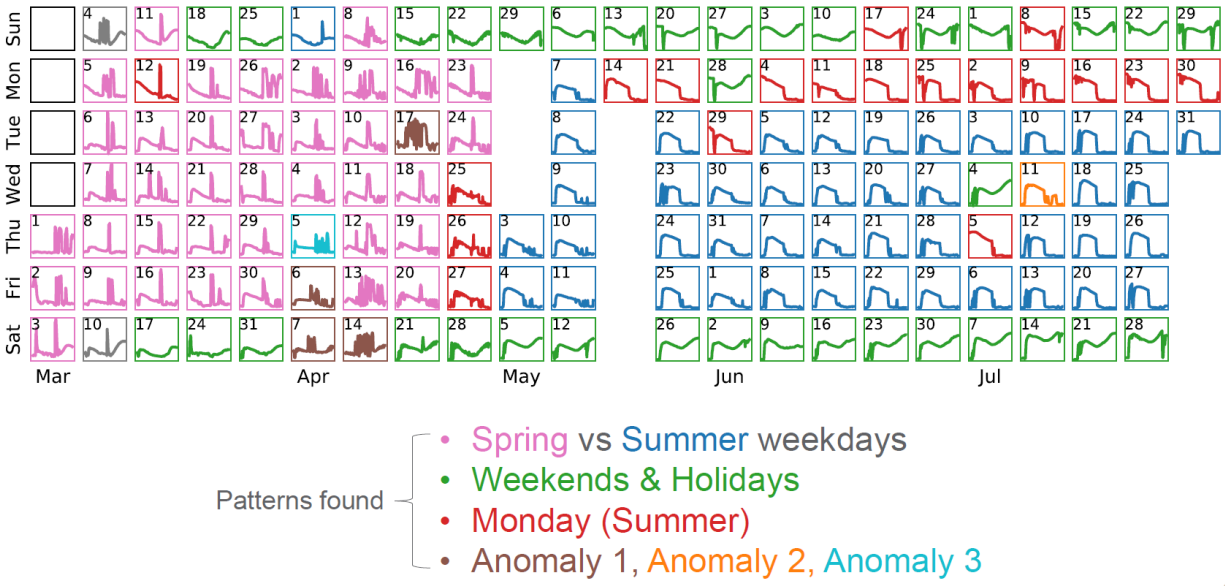


Figure 6 Groupings identified by CHISSL for building sensor data. Note that the July 4 holiday is grouped with weekends while July 5 matches a Monday

5.5 Initial Design for Information Framework

A behavior model (which is comprised of device identifiers and baseline communications and interactions) will be a critical component that informs the design of the potential Information Framework illustrated in Figure 6. Using this framework would determine the relationships between external and internal data flows through an organization's network. As we have a generic IoT network in IoTCOE, this will better inform how the flow will carry into other organizations. We will also show the relationships between the external network (cloud, vendor cloud) internal network systems, and subsystems.)

Design of the framework is in process and will be informed by further analysis of the network connections and communications collected. After gaining an understanding of this network interaction, we will then work up the stack to understand the software and hardware of the devices which generate this communication.

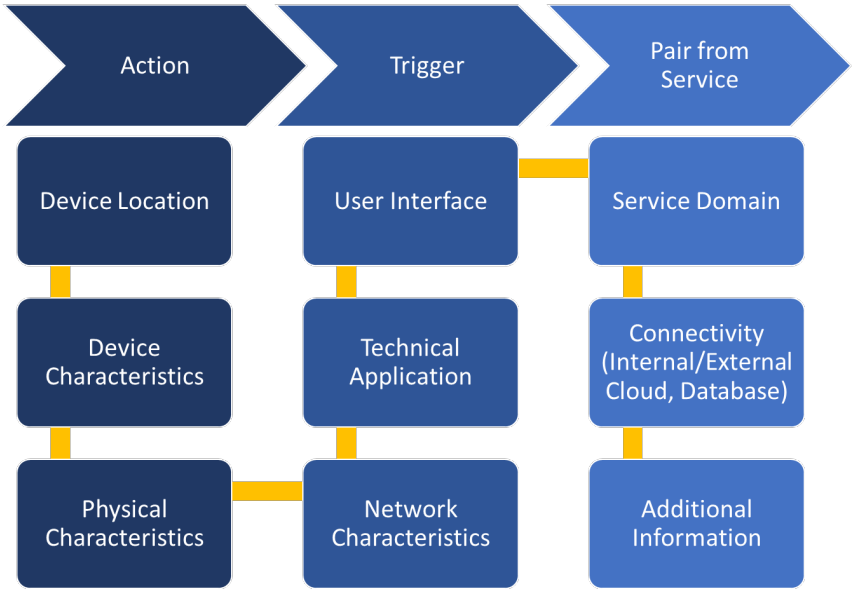


Figure 7 Initial Information Framework

Appendix A – Data Assessment Process

Detailed list of data assessment process

Order of operations for an assessment

- Ask the sponsor what kind of environment the device is in,
- what protocols will be enabled/disabled,
- what security will be used
- Can we build a testbed to discover these questions?
- Do a literature search: find the device and anything that talks to the device. Include software search, any third-party libraries that are compiled in, also known vulnerabilities on those.
- Assuming you have a test bed, watch the communications, look at how software normally works, how are users normally using it.

Map your attack surface, find out how devices are normally used/baseline.
Examine protocols, examine if they have known vulnerabilities, see if it is exploitable.
If there are known vulnerabilities, you can attempt to verify that they still exist in the device. You may or may not be able to duplicate a known attack.

Begin trying to find new vulnerabilities

- Blind Black box fuzzing. How is input entering the device? See if you can make the device misbehave, maybe a DOS?
- If you are able, begin doing grey box fuzzing. Look at static assembly versions of the code finding targets you may be able to attack. Easier to do with Dynamic code analysis.
- Dynamic code analysis used by developers of the system, is basically a debugger. Which enables debugging mode and you can step through the execution of the code on the device.

The cloud aspect of IOT may be a new challenge that we have not dealt with before. Something to look further into.
We can in fact reuse existing firmware. Which means if you get an older device in an established product line, you can look for code nobody uses anymore but is still there and may have vulnerabilities.

Appendix B – Data Points

Table 1 shows the network and device identifiers (i.e. data points) that the team used in the data captures and assessment process.

Table 1 Network Device Identifiers

Time Stamp	Sources	Network ID	MAC	IP	Destination IP	Host name
Destination Port	Port Source	Ethernet	Wireless	Bluetooth	Model #	Serial #
FCC #						

Appendix C – Software Data Captures

The figures in this section are snapshots of the information gathered from the various tools used during the Tardis study. These tools were very instrumental in the process for gathering information about the IoT devices chosen for the study. Further analysis shows that communications are complex, and each tool can provide additional insight in the characterization of the network device communications.

1. Kismet example of the data results

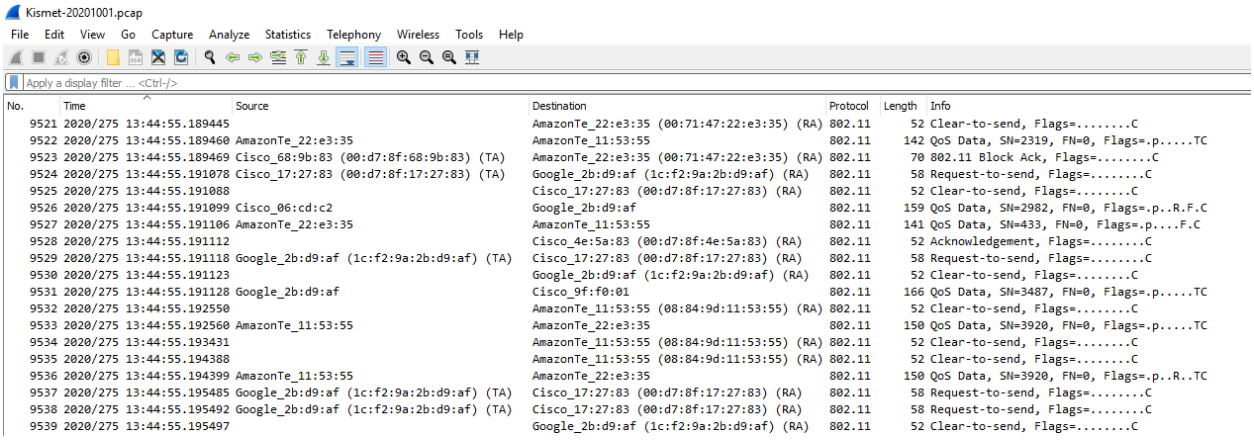


Figure 8 Kismet Bluetooth Captures

2. TCPDump example of the data results

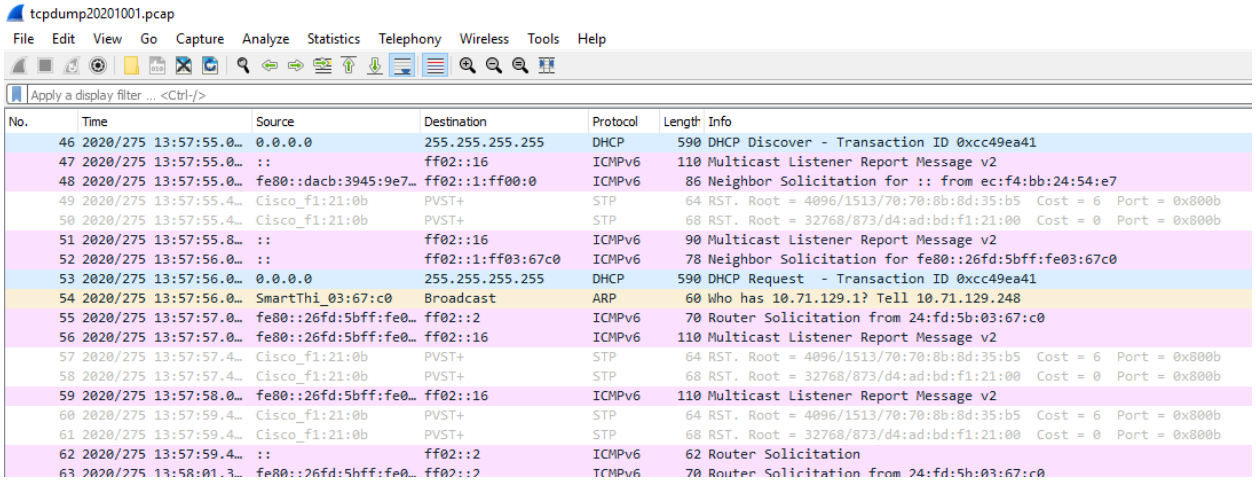


Figure 9 TCPDump

3. Wireshark example of the data results.

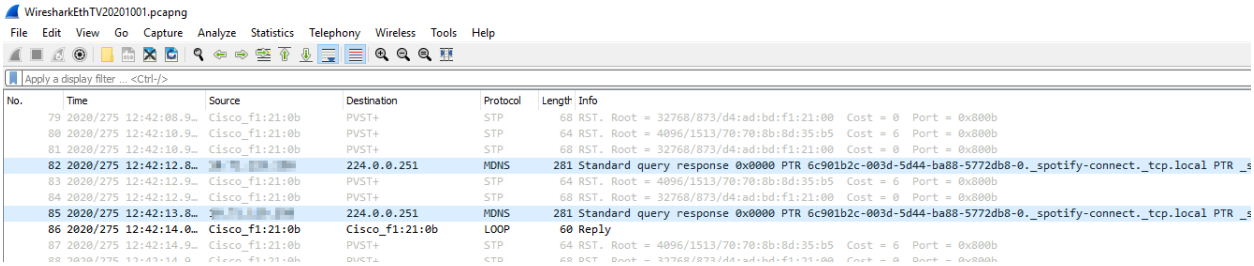


Figure 10 Wireshark

4. Colasoft Capsa an example of the data results.

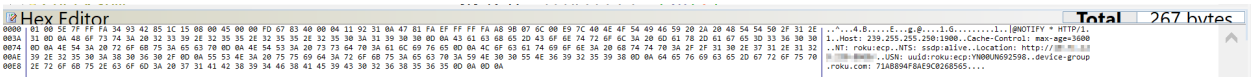


Figure 11 Colasoft Capsa

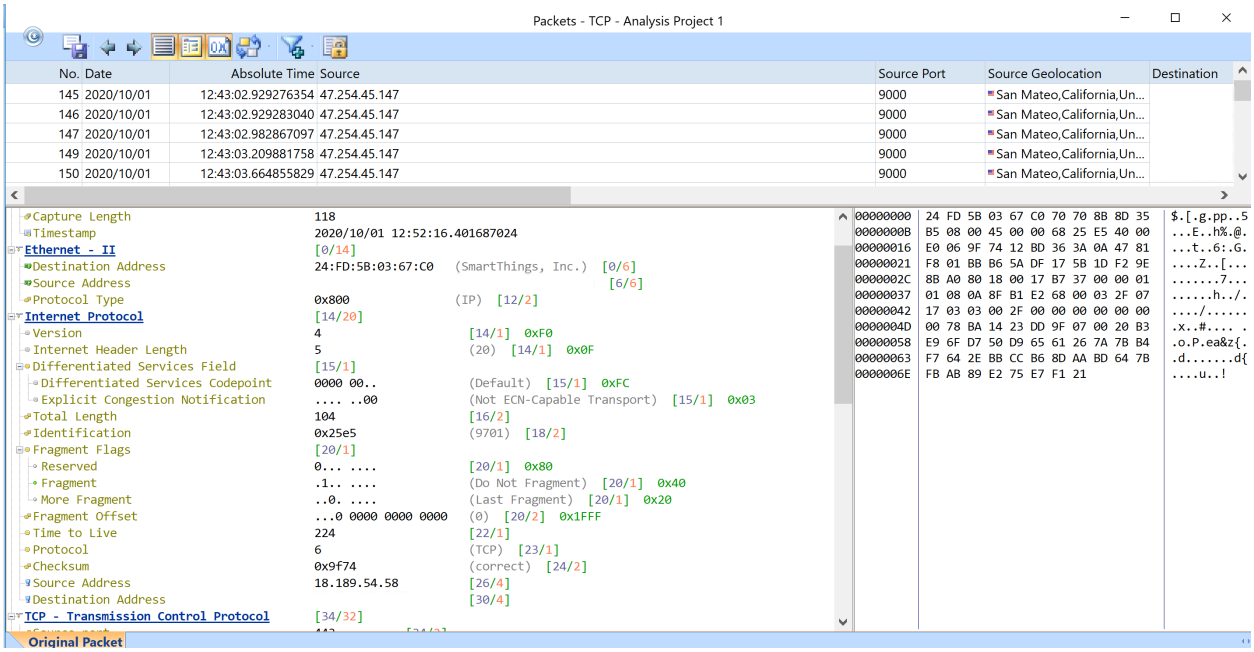


Figure 12 Colasoft Capsa Roku TV

5. Splunk example captures of the data results.

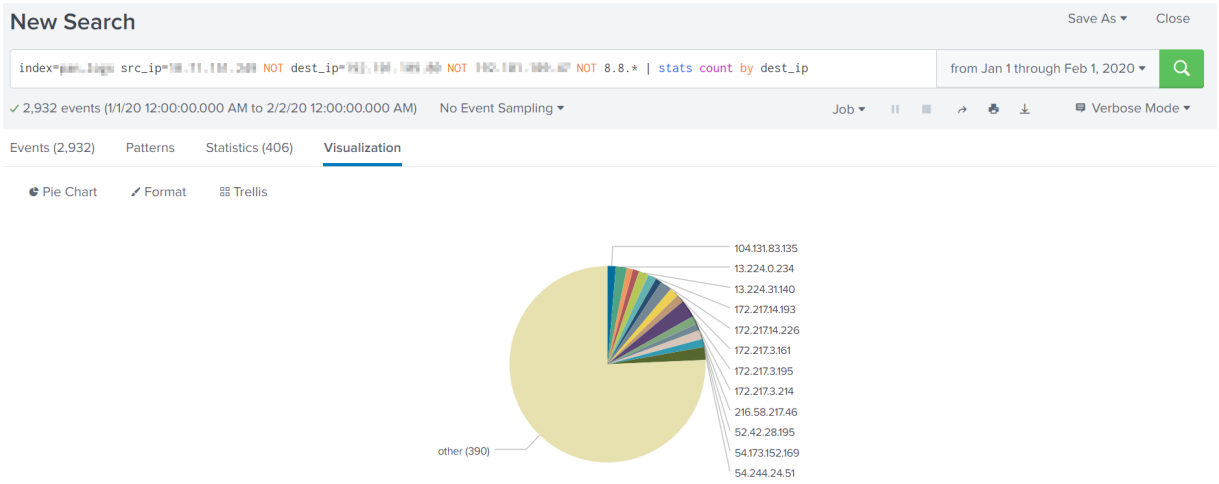
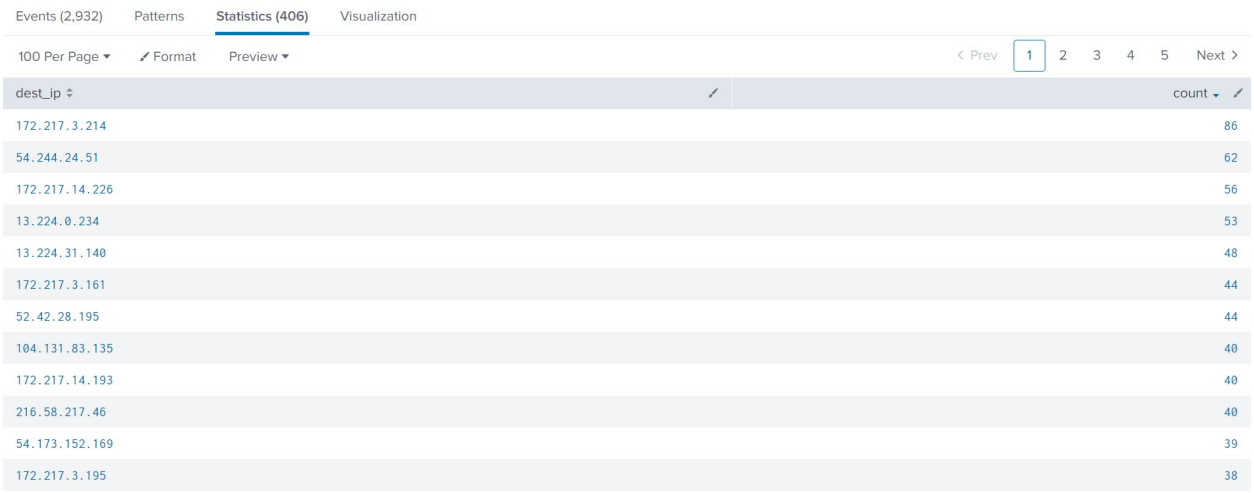


Figure 13 Splunk showing graphically the number of destination IP addresses with which the Roku TV communicated from January 1, 2020 to February 1, 2020



dest_ip	count
172.217.3.214	86
54.244.24.51	62
172.217.14.226	56
13.224.0.234	53
13.224.31.140	48
172.217.3.161	44
52.42.28.195	44
104.131.83.135	40
172.217.14.193	40
216.58.217.46	40
54.173.152.169	39
172.217.3.195	38

Figure 14 Destination IP addresses the Roku TV communicated with from January 1, 2020 to February 1, 2020

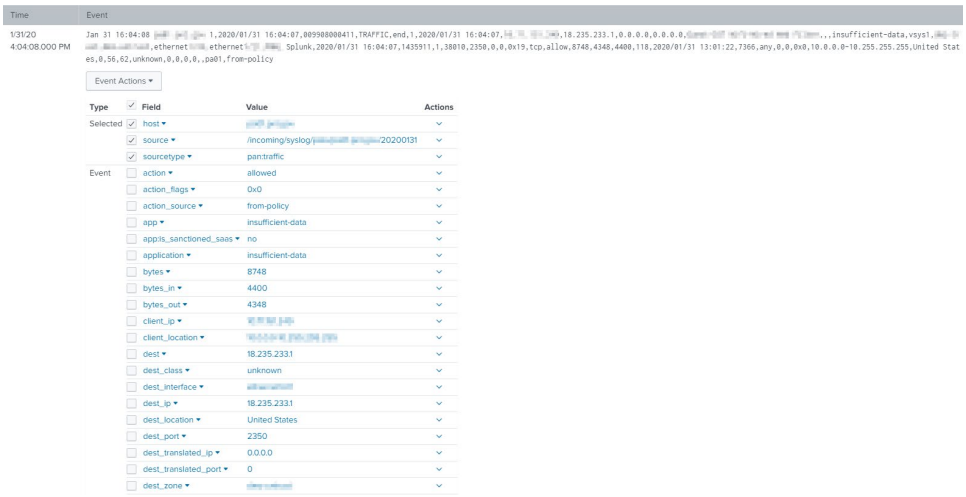


Figure 15 Splunk Data Fields

Figure 16 Splunk Data Fields

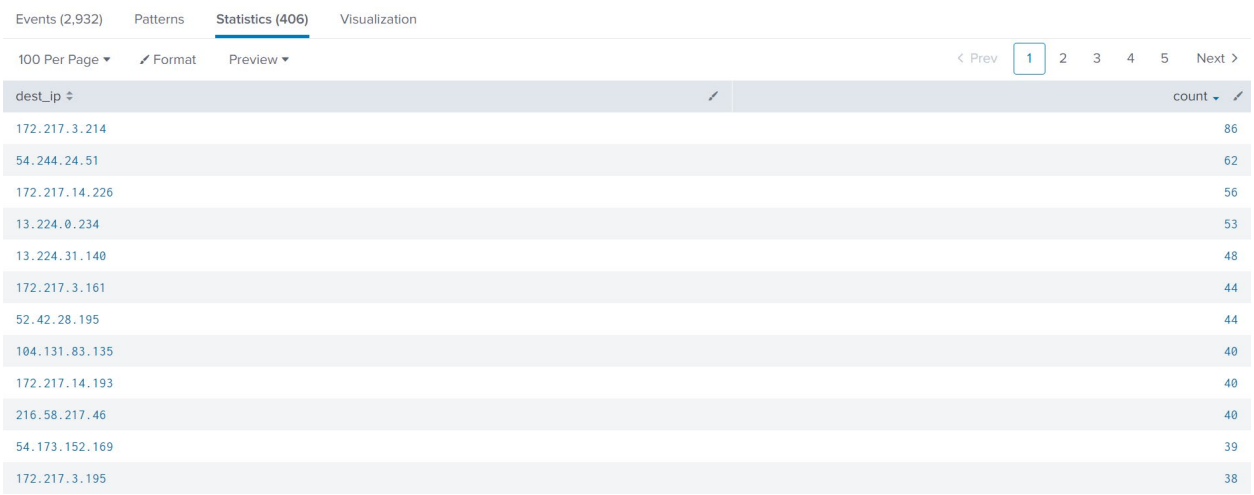


Figure 17 Splunk Network Device IPs

Appendix D Analysis Questions

1. How does the device communicate with the manufacturer/provider?
    a. Cloud
    b. Manufacturer
    c. Internal
2. What if the main function of the application/device?
    a. What does the data look like?
    b. What information is being transmitted?
    c. What does the connection like?
    d. What does the connection process look like?
    e. How does the device handle multiple connections?
    f. Does the behavior change if connected to a bigger network?
3. What is the identifier look like for the device?
    a. What is the purpose of this identifier?
    b. Does it have unique characteristics to the device?
4. How do you classify the identifier for the device?
    a. Is it a communication classifier?
    b. MAC address transmission
    c. Connection process
    d. Protocol
5. How does the classifier fulfill the uniqueness of the device?
    a. Scalability
    b. Privacy
    c. Data sets
    d. Communication
    e. Connection

Appendix E Identified Common Vulnerabilities and Exposures (CVE)

CVE for Roku TV

CVE-2018-11314	The External Control API in Roku and Roku TV products allow unauthorized access via a DNS Rebind attack. This can result in remote device control and privileged device and network information to be exfiltrated by an attacker.
CVE-2018-13989	Grundig Smart Inter@ctive TV 3.0 devices allow CSRF attacks via a POST request to TCP port 8085 containing a predictable ID value, as demonstrated by a /sendrcpackage?keyid=-2544&keysymbol=-4081 request to shut off the device.

CVE for the Nest Yale Keyless Lock

CVE-2019-17627	The Yale Bluetooth Key application for mobile devices allows unauthorized unlock actions by sniffing Bluetooth Low Energy (BLE) traffic during one authorized unlock action, and then calculating the authentication key via simple computations on the hex digits of a valid authentication request. This affects the Yale ZEN-R lock and unspecified other locks.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov