MCDERMOTT, C.D., JEANNELLE, B. and ISAACS, J.P. 2019. Towards a conversational agent for threat detection in the internet of things. In Proceedings of the *2019 International Cyber science on cyber situational awareness, data analytics and assessment (Cyber SA): pioneering research and innovation in cyber situational awareness, 3-4 June 2019, Oxford, UK*. Piscataway: IEEE [online], chapter 6. Available from: https://doi.org/10.1109/CyberSA.2019.8899580

Towards a conversational agent for threat detection in the internet of things.

MCDERMOTT, C.D., JEANNELLE, B., ISAACS, J.P.

2019

Towards a Conversational Agent for Threat Detection in the Internet of Things

Christopher D. McDermott, Bastien Jeannelle, John P. Isaacs

School of Computing Science and Digital Media
Robert Gordon University
Aberdeen, United Kingdom
Emails: {c.d.mcdermott, b.jeanelle, j.p.isaacs}@rgu.ac.uk

***Abstract*—A conversational agent to detect anomalous traffic in consumer IoT networks is presented. The agent accepts two inputs in the form of user speech received by Amazon Alexa enabled devices, and classified IDS logs stored in a DynamoDB Table. Aural analysis is used to query the database of network traffic, and respond accordingly. In doing so, this paper presents a solution to the problem of making consumers situationally aware when their IoT devices are infected, and anomalous traffic has been detected. The proposed conversational agent addresses the issue of how to present network information to non-technical users, for better comprehension, and improves awareness of threats derived from the *mirai* botnet malware.**

***Index Terms*—Situational Awareness, Intrusion Detection, Botnet, DDoS, Amazon Echo, Alexa, Virtual Assistant, Conversational Agent.**

I. Introduction

The Internet of Things (IoT) continues to grow and permeate into many areas of everyday life. Three areas of particular growth are in health, industrial applications and smart cities. Central to the future of smart cities, is the smart home, where an uptake of low cost and ease to deploy IoT devices, has already been witnessed. This flourishing smart home IoT market is fuelled largely by the promise of convenience, greater inter-connectivity and automation of everyday tasks [1]. As a result, smart devices such as IP cameras, doorbells with alert notifications, and light bulbs capable of being switched on using a digital assistant such as an Amazon echo, are increasingly becoming commonplace in the home. Whilst smart interconnected devices clearly have many benefits, concerns still exist around the security and privacy of such devices, and data derived therein. A trend is evident whereby market forces, and the need to be competitive, have placed IoT manufacturers under increasing pressure to produce low cost, plug and play smart devices. Popular with consumers, these devices often omit vital security and privacy mechanisms (to promote simplicity and adoption), exposing devices to potential threats and leaving them vulnerable to potential attackers. Arguably one of the most serious threats facing IoT devices, is that of botnets. The vast threat landscape afforded by the IoT, and the inherent vulnerabilities of many smart devices, has provided the perfect platform to perform large scale distributed denial of service (DDoS) attacks [2]. Indeed, many powerful DDoS attacks have been witnessed in recent years, with the most prominent example being the *Mirai* botnet, which denied service to some of the most widely used platforms on the Internet such as Twitter, Netflix and Reddit [2].

A common trait of many of these high profile DDoS attacks, has been their exploitation of smart devices commonly found in consumer homes, such as IP cameras, and home routers [3]. In our previous work [1], we investigated if a representative sample of consumers (n=158) were able to identify if an IoT device was infected with malware such as *mirai*, and was being used to perform DDoS attacks. A cross-sectional study was performed to allow multiple variables to be compared. To evaluate consumers perception and awareness of threats facing the IoT, the sandboxed botnet environment established in [3] was used. This consisted of a (C&C) server, to remotely control the infected smart devices (bots), and a scan/loader server to infect new smart devices with the botnet malware. Several smart IP cameras were infected, and used as bots to perform a variety of DDoS attacks, during which live video feeds were recorded. The consumers were presented with the recorded video feeds, and asked to identify which cameras were infected. Results demonstrated that across all age ranges (*18-60+*) and technical abilities (Novice-Expert) it was very difficult for consumers to be situationally aware of infected IoT devices.

This paper presents a solution towards the detection of botnet activity within consumer IoT devices and networks. A conversational agent was implemented, and its effectiveness to improve consumers situational awareness, tested. We implemented our approach by using our previously created dataset [3], which was amended and simplified for application in this study (detailed in Section III-A). An Amazon Echo device was used to interact with a developed Alexa Skill, which sent user queries to an AWS Lambda function, which in turn queried the dataset, and returned an appropriate response to the user. Thus, the main contributions of this paper are:

1) A scalable serverless ETL pipeline for parsing intrusion detection logs;
2) A novel conversational agent utilising aural analysis for detecting anomalous traffic in consumer IoT networks.

The rest of the paper is organised as follows: Section II introduces the concept of situational awareness, specifically it's application to the Cyber domain. It also presents common methods of network monitoring and threat detection. Finally, the use of virtual assistants within the IoT is discussed. Section III details the methodology used to generate the necessary classified network traffic, and means of querying the data by use of aural analysis. Implementation of the ETL pipeline and conversational agent are presented in Section IV. Section V describes participant recruitment, evaluation undertaken and results gained. Section VI presents final conclusions, study limitations, and future research suggestions.

II. Background

For our related works, we shall consider the topics Cyber Situational Awareness (SA), common methods of network monitoring and threat detection, and the growing use of virtual assistants in the IoT. Although other threats exist, the rise of botnet activity in the IoT, is considered of utmost importance, and will therefore form the basis of threats considered in this study.

A. Cyber Situational Awareness

Situational awareness (SA) is often traced back to the seminal work presented by Endsley in [4]. A study was presented which investigated if enhancing SA in aircraft pilots could increase their likelihood of making optimal decisions in dynamic situations. In [5] the author continued the work and presented a SA theoretical model, applicable across a variety of environments and systems, beyond aviation. Here, the author defines SA as a person's state of knowledge about a dynamic environment. Specifically, their perception of elements in the local environment, the comprehension of their meaning and relevance to the person's goals, and a projection of future states of the environment based on this understanding. The SA model presented by Endsley in [5] is considered of central importance to SA research, and has therefore been widely adopted as a reference model, and subsequently applied to a broad range of research areas. The model is comprised of three levels, namely *Perception*, *Comprehension* and *Projection* which combine and contribute to achieving a level of awareness in a given situation. They can be defined as:

1) *Perception*: the consciousness of relevant elements in the environment, specifically the status, attributes, and dynamics of elements in relation to the environment.
2) *Comprehension*: the synthesis of the seemingly disjointed elements at level 1, to understand their significance, fuse together to derive meaning and patterns, and foster a holistic understanding of the environment.
3) *Projection*: the ability to project the current situation of the environment into the future, predict the likely subsequent actions of elements, ultimately allowing better decisions to made in dynamic situations.

In [6] McGuiness and Foy extended Endley's model to include an additional level, defined as *resolution*. Here, the aim is to establish the best course of action to take to change the current situation to the desired state. Resolution is achieved by considering all possible actions from a range, and selecting the most appropriate course of action accordingly [7].

When applied to the Cyber domain, Cyber SA can be defined as the compilation, processing and fusing of network data to understand a network environment and accurately predict and respond to potential threats that might occur. Seminal work by Denning [8] focused on the detection of cyber attacks, leading to the Joint Directors Laboratories (JDL) creating a conceptual data fusion model which identified the processes, functions, categories, and specific techniques applicable to data fusion [9]. Drawing similarities to Endsley's model it defined levels for *Data Assessment*, *Object Assessment*, *Situation Assessment*, *Threat Refinement*, and *Process refinement*. Importantly, it highlighted the importance of human elements in achieving SA. In [7] Tadda combined the JDL Data Fusion model with Endsley's SA model to propose a Situational Awareness model applicable to the Cyber domain. The authors addressed the differences between level 2 and 3 of the JDL model and Endsley's Projection level. In doing so, they argued that a computer system is capable of identifying the occurrence of an activity based on priori knowledge and cannot itself develop or provide Situation Awareness; only a person (the decision maker) can derive the awareness. They drew comparisons between the two models and asserted that level 2 of the JDL model and Endsley's *Comprehension* level address the current situation. Whereas, level 3 of the JDL model and Endsley's *Projection* level address the ability to project the current situation into the future, in order to predict future impacts and threats. Essentially, they propose splitting level 2 and 3 JDL assessments based on time rather than functionality. Other prominent researchers in the Cyber domain have used these models, in particular Endsley's model, to further research in this area. In [10] Onwubiko identifies the functional attributes of situational awareness for network/cyber security. A SA model for network security is presented and ten fundamental attributes are suggested, which the author proposes should be considered when implementing any SA system in the domain. In [11] the author extended the work and presented an adapted version of Endsley's SA reference model [5]. The model incorporated Endsley's initial levels *Perception, Comprehension and Projection* and also the fourth level *Resolution* proposed by McGuiness and Foy [6]. The proposed Cyber SA Instantiation Model overlays Endsley's model but is generalised to be applicable across the Cyber domain. An additional fifth awareness level is presented and fuses with the previous four levels as follows:

L0 - Information Generating Sources: Log sources such as event logs, which are evidence of an attack or exploit, but are unable to detect an attack without functions from the subsequent levels.

L1 - Perceive: use of individual toolkits to gather raw data from Level 0 about perceived situations in the network. Information is classified into meaningful representations to form the basis for comprehension. Four distinct sources of information

are identified which contribute to this level namely, *Protection sources*, *Threat Intelligence sources*, *Tracking sources*, *External Intel sources*.

L2 - Comprehend: use of analysis tools and techniques to continually anaylse and synthesise information from Level 1. Fusions of disparate events and correlation of information from multiple sources, to link evidence and gain an holistic overview of the situation.

L3 - Projection: analysed intelligence once comprehended, can be used to predict future events and situations. Performed as a real-time continuous process, allows possible mitigations against threats to be recommended.

L4 - Resolve: recover and resolve situations using mitigation strategies identified in level 3. Coordination is required for triage, investigation, classification, and prioritisation in order to resolve, remedy, and recover events and Cyber situations.

SA when applied to the Cyber domain is still relatively immature as a research area. The general models discussed here, and adapted versions for the Cyber domain, do however form a good basis for assessing and enabling the application of SA in the Cyber domain.

B. Network Monitoring and Threat Detection

Network monitoring and Threat detection are common tasks undertaken to foster greater SA of activity on a network. Administrators collect logs from a variety of sources, such as Intrusion Detection/Prevention Systems or Traffic Analysis Software, to understand baseline behaviour, and identify anomolous traffic on the network. When using alerts from an IDS to monitor a network, systems can be classified by two distinct methods, with a third hybrid approach also available.

Signature-based Detection. Uses known rules (signatures) of previous attacks, and compares these against collected data. Matches found invoke an alert, with high true positive and low false positive rates possible for known attacks. Conversely, signature-based detection can return low detection rates for zero day (new) attacks [12]. *Anomaly-based detection*. Compares current network traffic with a baseline of normal network behavior. Any significant deviation from this baseline is detected and classified as an anomaly, raising an alert. Good detection for new or unknown attacks is often achievable, although high false positive rates are also common [12]. *Specification-based Detection*. A hybrid of the previous two methods, where specifications are developed to describe normal network behavior. Two detection mechanisms are usually combined, one to detect known attacks using signatures, the other to monitor traffic and detect deviations from normal network behaviour. Low false positive rates can often be achieved, compared with just anomaly-based detection methods [12].

Irrespective of the monitoring tool used, data is commonly conveyed to the administrator through a variety of methods, including data visualisation. In [13] the authors investigated visual analytic methods for log files, and concluded this to be an effective way for humans to identify patterns in traffic. This view is corroborated in other studies [14] [15], where authors in [16] suggest the use of visualisation improves situational awareness, since it aids in perceiving and comprehending the current status of the network, with prediction of future situations also possible.

Despite promise, some authors highlight shortcomings in text and visual based monitoring systems, arguing that they require the full attention of the administrator to prevent missing key information [17]. They propose the use of data sonification, as a method to improve situational awareness of network activity. Here, raw data is presented in audio form (generally non-speech), and has been used across a wide range of fields. It's application to computer security is summarised in [18], where the author suggests that due to its time based nature, and our auditory cognition, sonification is especially suitable for data that changes over time.

In [19] the authors present a formalised model for designing sonifications for network security monitoring. They conducted a study in [20] investigating the attitudes towards using sonification by security practitioners. Results showed high potential for its use in peripheral monitoring (whilst undertaking other tasks), demonstrating it could be combined with visualisation to address its limitations, and form a multimodal approach to network monitoring.

C. Use of Virtual Assistants in the IoT

A Virtual Assistant (VA) is a term often used to describe a spoken dialogue system, which uses an intelligent agent to assist users to complete tasks through auditory interactions. Growing in popularity, they have been widely adopted by a range of companies, producing Microsoft's *Cortana*, Apple's *Siri*, Google's *Assistant*, and arguably the most popular, Amazon's *Alexa*. Devices such as Amazon's *Echo* and its conversational agent *Alexa*, provide opportunities to build feature rich conversational interactions.

Research in this area is growing, and producing some very promising applications of virtual assistants. In [21] a system is presented using Amazon's *Alexa* and *Node-Red*, a simple and powerful automation platform, to interconnect and control numerous IoT devices. The system provides the ability to switch smart devices e.g. lightbulbs on/off, monitor iPhone statistics, and use voice commands to control a heater. The research offers a lot of promise in this area, but a lack of details made it difficult to fully assess functionality.

In [22] a smart home system is presented, using the *Reverb* and *Telegram* mobile apps, to control smart appliances in the home. The reverb app is used to send voice commands to the Alexa Voice Service in AWS, which interfaces with a local raspberry pi, to switch a device on/off. The Telegram app is used to send commands via text, and perform similar tasks. Functionality was limited, but returned positive results, demonstrating good promise in this area.

The next generation of virtual assistants is presented in [23]. A multi-modal dialogue system is developed to combine multiple user input modes, such as speech, touch and verbal/non verbal gestures. The authors propose a system which uses a camera and kinnect device to receive speech and gesture input commands, which are processed and stored in

a knowledgebase. The application of the system was unclear, however future use cases are suggested including, educational assistance, robotics and home automation.

In [24] the use of virtual assistants to assist the elderly, was assessed. The authors propose that virtual assistants could be used to combat social isolation amongst elderly people. Microsoft's *Cortana*, Apple's *Siri*, Google's *Assistant*, and Amazon's *Alexa*, were tested for their ability to complete tasks, which could improve the issue. Each assistant was tested for functionality, and their ability to provide a *basic greeting*, *email management*, *social media*, and *social games*. The results presented were inconclusive, but did demonstrate the range of applications, and problems, virtual assistants could be used to address.

Finally, some interesting research has been conducted in [25] which looked at the personification of virtual assistants, such as the Amazon Echo. They found 30% of customers would like to treat the Amazon Echo as a human character due to its personified name (*Alexa*) and ability to talk. It is clear from the research presented, that virtual assistants offer a wide range of use cases and applications. The willingness of users to adopt this new method of interacting with devices and information, and trust towards them, could promote wider use in the future.

Having reviewed the existing literature, a gap is identified regarding the use of aural analysis for threat detection and network monitoring, specifically the use of virtual assistants to aid situational awareness. Research in [19] [20] demonstrated the promise of sonification in the area, however studies appear to focus on non-speech methods. The aim of this paper, is to make a contribution in this area, specifically using a virtual assistant to provide aural analysis using Natural Processing Language (NLP) methods.

III. Methodology

An overview of the methodology used to generate the necessary classified network traffic, and means of querying the data through aural analysis, is presented below.

A. Data Sources

In our previous work [3], a secure sandboxed environment was created, and a dataset containing IoT botnet traffic was generated. The generated dataset consisted of 37 captures (3600 second duration each), over a total of five days, and was stored in *pcap* and *csv* format. Ground truth labels were assigned, and subsets of the dataset were used to test intrusion detection models in [3]. The full dataset is available upon request.

To test the conversational agent presented in Section III-C, a subset of this dataset was used in this study, containing both background *(classified as normal)* and IoT botnet related traffic *(classified as unusual)*. To aid better understanding of the data, features were renamed from *No. Time*, *Source*, *Destination*, *Protocol*, *Label* to *ID*, *DateTime*, *SourceDevice*, *DestinationDevice*, *DataType*, *Activity*. Although features *Length* and *Info* were used during the detection and classification of threats in [3], the complexity of the information meant they had limited value for use in the conversational agent. Since they would not be required later, they were removed.

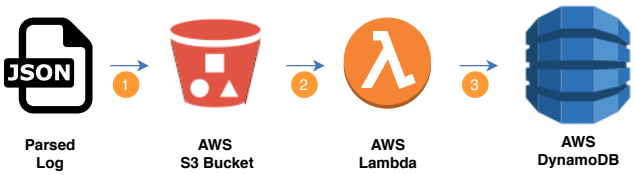

Fig. 1. IDS Log Parsing and Storage

Finally, the *csv* files were concatenated, converted to *JSON* format, and stored in a specified directory ready to be ingested by the ETL pipeline described in Section III-B. A sample record from the newly amended dataset is found in Source Code 1.

```
{
  "ID": "487",
  "DateTime": "20/01/2019 19:01",
  "SourceDevice": "192.168.252.40",
  "DestinationDevice": "180.130.236.179",
  "DataType": "TCP",
  "Activity": "normal"
}
```

Source Code 1. Sample JSON record

B. ETL Pipeline for Parsing IDS Logs

The ETL pipeline for handling and uploading IDS logs is presented in Fig. 1. Suitable IDS logs can be parsed and stored in the specified directory, ready to be ingested by the ETL pipeline. For our study, the IDS logs consisted of the amended dataset described in Section III-A. In *(step 1)* a script monitors the directory for new files. When a new *JSON* file is added, the file is extracted, transformed, and loaded to an *S3* bucket on AWS. In *(step 2)* a Lambda function is triggered whenever a new file is added to the *S3* bucket. The use of Lambda allows code to be executed without provisioning or managing a server. It also ensures costs are reduced since they only occur when a function is triggered, and code run. In *(step 3)* once the *handler object* has been triggered, the code in the Lambda function is executed, and data loaded into the *DynamoDB* Table.

C. Conversational Agent Architecture

Primary input for the conversational agent is speech derived from Amazon Alexa enabled devices. Input is analysed using natural language processing (NLP) techniques to understand the user query. Requests are then matched against the secondary input source (IDS logs stored in a DynamoDB table), and responses are returned accordingly.

The agent consists of three main components: a database of classified IoT traffic, NLP engine as an interface between a user and the Alexa device, and a query handler. In the presented conversational agent, the speech recognition engine is contained in the Alexa device, the query handler is our developed Alexa Skill and AWS Lambda function, and a

DynamoDB database are used to store and query classified IDS logs.

In Fig. 2, the agent frontend is powered by an Amazon Echo device. NLP software in the Echo device uses speech recognition to convert user input (in the form of speech), to text. The query handler acts as the bridge between the Echo device and the IDS database. The Alexa skill receives converted aural requests from the Echo device *(step 1)*, and forwards the request to the AWS Lambda function *(step 2)*. A query request to interact with the DynamoDB table is triggered *(step 3)*, which when fulfilled returns an appropriate answer to the user query. Finally, the the Alexa skill generates an aural response from the returned answer, invokes the Echo device, which communicates an aural repsonse to the user *(step 4)*.

The backend of the system is hosted on AWS infrastructure as a scalable serverless solution, which parses and stores IDS logs in a DynamoDB table. The handler function is hosted on AWS Lambda, which is a server-less technology that allows event-driven code to be run without provisioning servers. The handler function is used to trigger interaction with the DynamoDB, and provide functionality to the Alexa Skill.

IV. Implementation

To promote reproducibility of this paper, a detailed description of the ETL pipeline and conversational agent are presented below.

A. ETL Pipeline

The implementation of the ETL pipeline required three processes, as shown in Fig. 1. In *(step 1) crontab* was configured on a local *raspberry pi* to run a script on a specified schedule. The script monitored a local directory for new *IDS* logs, and invoked a process to upload newly added *JSON* files to an *S3* bucket on AWS. For our study, the amended dataset from Section III-A was manually added to the directory and processed by the ETL pipeline.

To handle the backend functionality, an *IAM Role* was required. From the AWS Management Console, a new *IAM Role* was created and (*AmazonS3FullAccess*, *AmazonDynamoDB-FullAccess*, and *AWSOpsWorksCloudWatchLogs*) permissions assigned.

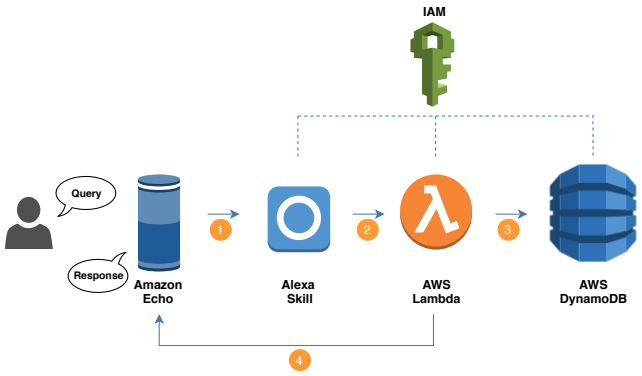

Fig. 2. Conversational Agent Architecture

In our pipeline, a Lambda Function was used to transfer items from the *S3* bucket into DynamoDB. First, a table was created using the attributes found in the *JSON* dataset (see Source Code 1). DynamoDB is a schema-less database that only requires a table name and primary key. The tables primary key is made up of one or two attributes that uniquely identify items, partition the data, and sort data within each partition. The *ID* attribute was set as the primary key to uniquely identify items.

The Lambda function was configured to be triggered when a file upload event occurs in the configured S3 bucket. *Lambda_handler (event,context)* was configured as the handler to start the AWS Lambda function. Once called, the function was configured to wait for data to be retrieved through the S3 service, before reading the *JSON* file. Data was then passed to the *insert_data()* function, which takes control of the table first, then iterates through the list and inserts it into the table using the *put_item* function.

B. Frontend Agent Architecture

From the Alexa Developer Console, a new skill named *Threat Detector* was created. An invocation name was assigned, and is used to invoke the Alexa Skill from the Echo device. Twelve intents were configured, and used to trigger specific event functionality. Seven in-built intents were used as triggers to perform preconfigured functionality such as *repeat*, *stop* or *cancel* an intent. Five custom intents were configured to enable a user to query the DynamoDB table for information, as detailed below:

1) ***activitySummaryToday:*** Responds to a user query and returns a summary of all activity taking place today.
2) ***activitySummaryByDate:*** Responds to a user query and returns a summary of all activity taking place on a specified date.
3) ***activitySummarySrcDevAndDate:*** Responds to a user query and returns a summary of all activity from a specified source device on a specified date.
4) ***firstUnusualActivityByDate:*** Responds to a user query and returns details of the first activity on a specified date, which is classified as unusual.
5) ***activityDetailsByID:*** Responds to a user query and returns details of a specified activity ID.

For each custom intent a series of *utterances* were configured. Utterances are the phrases a user may use to trigger a particular intent. Given the variation of spoken language in the real world, there will often be several ways to express the same request. To invoke the *activitySummaryToday* intent a user could say “show me a summary of today’s activity”, “show me the summary of today’s activity ” or “show me summary for today’s activity ”. To ensure an intent could be invoked using a variety of expressions, a minimum of three sample utterances were configured for each custom intent.

Utterances which contained words that represent variable information a user will specify, were assigned a *slot*. For example, to invoke intent *activityDetailsByID* the utterance “show me details for activity id {ID}” was used, where the

{ID} slot would be replaced with an id number specified by the user, such as *three hundred sixty six*.

Finally, the endpoint is set to *AWS Lambda*, since the Alexa Skill will invoke the Lambda function to process the identified request and return a response which is spoken back to the user.

C. Backend Agent Architecture

The main components of the backend architecture are the *AWS Lambda* function and *DynamoDB* table. To control access to backend resources, the *Identity and Access Management(IAM)* service was used to control authentication and authorisation. An *IAM Role* was created and inline policies assigned for DynamoDB access and AWS Lambda execution, to allow the Alexa Skill to invoke the Lambda function as its backend.

A DynamoDB table was created using the attributes found in the JSON dataset (see Source Code 1).

The main engine of the backend query handler, is the AWS Lambda function. From the AWS Management Console, a new Lambda function was created, runtime environment specified, and previously created *IAM* role attached. A *handler object* was specified, which serves as the hook that AWS Lambda uses to execute the code in the Lambda function. *Alexa Skill Kit* was specified as the trigger to execute the Lambda function, and the Alexa Skill *ID* was input as the endpoint to receive POST requests when a user interacts with the Alexa Skill. Finally, to link the Lambda function to the Alexa Skill, the *Amazon Resource Number (ARN)* of the Lambda function was set as the endpoint for the Alexa Skill, in the Alexa Developer Console.

V. Evaluation and Discussion

TABLE I. Assessment of Situational Awareness across Use-Cases

	Pre	**Post**			
Statement	*Mo**	*Mo*	*Md***	*W*	*p*
s1	2	4	2	78	0.003
s2	2	3	1	36	0.014
s3	1	1	0.5	28	0.022
s4	4	4	0	1.5	0.999
s5	4	4	0	2	0.999

** Mode* *** Median*

A. Participants

The Universitys Research Ethics Policy was followed to establish and promote good ethical practice in the undertaking of this study. In doing so, consent to participate was implied when participants decided to engage in the research and complete the agent evaluation. Convenience sampling was employed, with subjects selected due to their convenient accessibility, and proximity to the author. A representative sample (n=12) of users completed the agent evaluation. Results of the agent evaluation are presented in Section V-C.

B. Use-Case Development

To demonstrate the capabilities of the conversational agent, five use cases were identified, mapped to the custom intents, and used to test the functionality of the conversational agent.

Use-Case One: In this scenario, a user is not aware of any threats or unusual activity within the network, and would simply like to request a summary of all activity taking place today. The user invokes the Alexa Skill using the specified invocation name (*Threat Detector*), and then proceeds to request a summary of activity. If the input is matched by the speech recognition engine to one of the specified custom intents and utterances (described in Section IV-B), a query is triggered, and a response returned. Below is the conversation as a result of the user query:

> ***User:*** "Alexa open Threat Detector"
>
> ***Alexa:*** "Hi, welcome to Threat Detector, how can I help ?"
>
> ***User:*** "show me a summary of today's activity"
>
> ***Alexa:*** "Sure, Today there has been 100.0 percent normal activity and 0.0 percent unusual activity. Is there anything else I can help you with ?"

Use-Case Two: In the second scenario, a user suspects a threat or unusual activity has occurred within the network. They invoked the intent associated with the first scenario and have been told there has been no unusual activity today. The user proceeds to query the database for unusual activity on a different date (perhaps yesterday). Again, if the input is matched by the speech recognition engine, to one of the specified custom intents and utterances, a query is triggered, and a response returned. Below is the conversation as a result of the user query:

> ***User:*** "show me a summary of activity for twenty january two thousand nineteen"
>
> ***Alexa:*** "No problem, On 2019-01-20 there was 94.74 percent normal activity and 5.26 percent unusual activity. Are there any other activities you would like to know about ?"

Use-Case Three: In the third scenario, a user suspects a threat or unusual activity has occurred within the network, on a specified date. They suspect a specific device may be causing the problem, so proceed to query the database for unusual activity on a specified date, by the given source device. Again, if the input is matched by the speech recognition engine, to one of the specified custom intents and utterances, a query is triggered, and a response returned. Below is the conversation as a result of the user query:

> ***User:*** "show me a summary of activity by one nine two dot one six eight dot two five two dot forty on twenty january two thousand nineteen"
>
> ***Alexa:*** "Of course, Summary for 2019-01-20 For device 192.168.252.40 there was 94.74 percent normal activity and 5.26 percent unusual activity. How can I be of further assistance ?"

Use-Case Four: In the fourth scenario, a user suspects a threat or unusual activity has occurred within the network, on a specified date. They are unsure which device has caused the problem, so proceed to query the database for the first unusual activity on the specified date. The intention is to identify which specific device may be causing the problem. Again, if the input is matched by the speech recognition engine, to one of the specified custom intents and utterances, a query is triggered, and a response returned. Below is the conversation as a result of the user query:

> ***User:*** "show me details of first unusual activity on twenty january two thousand nineteen"
>
> ***Alexa:*** "My pleasure, First unusual activity on 2019-01-20 For which, I.D is 410 Date Time is 20/01/2019 19:01 Source Device is 192.168.252.40 Destination Device is 118.56.238.213 Data Type is TCP Activity type is unusual. Is there anything else I can help you with ?"

Use-Case Five: In the fifth scenario, a user suspects a threat or unusual activity has occurred within the network, on a specified date. They have invoked other intents to query the database, and now suspect they know when a threat or unusual activity took place. They now query the database to get full details of the activity ID of when the unusual activity occurred. Again, if the input is matched by the speech recognition engine, to one of the specified custom intents and utterances, a query is triggered, and a response returned. Below is the conversation as a result of the user query:

> ***User:*** "show me details for activity id three hundred sixty six"
>
> ***Alexa:*** "I.D is 366 Date Time for that is 18/01/2019 19:01 Source Device is 192.168.252.21 Destination Device is 192.168.252.60 Data Type is TCP Activity type is normal. What else would you like to know?"

C. Agent Evaluation

To evaluate our conversational agent, each participant was asked to complete a pre-test survey and indicate their level of agreement with five statements, using a Likert-type scale from *Strongly Disagree* to *Strongly Agree*.

> ***s1:*** *I am confident I can tell if my home network is functioning normally.*
>
> ***s2:*** *I am confident I can tell if an IoT device on my network has been compromised.*
>
> ***s3:*** *I am confident I can tell if an attack has taken place on my home network.*
>
> ***s4:*** *I am confident I can tell which IoT devices are using my home network.*
>
> ***s5:*** *I am confident I can tell if an IoT device is functioning normally.*

Subsequently, participants were asked to use the conversational agent for the five use-cases presented in Section V-B and record their answers to the queries. Finally, with the new tool available to them, participants completed a post-test survey of the same five statements, and variance in their attitudes was recorded (see Table I). Since we were comparing related groups, where participants completed the same survey pre and post test, Wilcoxon signed-rank tests were used for comparisons. Statements (*s1-s3*) demonstrated a statistically significant median increase in agreement level, between *pre* and *post* test surveys. Statements (*s4-s5*) did not demonstrate any variance. Pre-test mode values for statements (*s1-s2*) demonstrated participants disagreed with the statements, suggesting participants were not initially confident they could tell if their network was functioning correctly or if a device had been compromised. The Post-test median increase would suggest the use of the conversational agent had improved situational awareness in these areas. To a lesser extent this was also true for statement (*s3*). Statements (*s4-s5*) did not show any median increase, however the Pre-test mode values for these statements suggested the participants were already confident they knew which IoT devices were using the network, and functioning correctly.

D. Suggested Improvements

On conclusion of the post-test survey, participants were asked for suggestions of possible improvements to the conversational agent. For brevity, these are summarised as follows:

1) ***normal network functionality:*** The ability to get a simple status of the network and if any unusual activity has occured.
2) ***compromised device:*** The ability to see which devices have been active on the network on a given date, and their total activity.
3) ***prescence of an attack:*** The ability to see the total activity for a device, and combined total for the network.

VI. Conclusions and Future Work

This paper presents the implementation of a novel conversational agent for detecting anomalous traffic in consumer IoT networks. In Section I, we presented the problem of detecting threats within consumer IoT networks. We demonstrated that without any clear signs of infection, it was very difficult for consumers to know when their devices are part of a botnet, performing large scale DDoS attacks. Results in Section V, clearly demonstrate that the presented agent could make a positive contribution towards improving situational awareness of threats in IoT networks. To the best of our knowledge, this is the first study to use a conversational agent to perform aural analysis in this domain. Despite this, limitations were identified in the study. Push notifications could further improve SA, however are not currently permitted by AWS. In the future, we plan to increase the functionality of the agent to include the suggested improvements highlighted by participants. In addition, we intend to investigate the suggestion in [20], and evaluate a multimodal approach to threat detection.

References

[1] C. D. McDermott, J. P. Isaacs, and A. V. Petrovski, "Evaluating awareness and perception of botnet activity within consumer internet-of-things (iot) networks," *Informatics*, vol. 6, no. 1, 2019.

[2] H. Sinanovi and S. Mrdovic, "Analysis of mirai malicious software," in *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Sept 2017, pp. 1–5.

[3] C. D. McDermott, W. Haynes, and A. V. Petrovski, "Threat detection and analysis in the internet of things using deep packet inspection," *International Journal on Cyber Situational Awareness*, vol. 3, no. 1, pp. 61–83, 2018.

[4] M. R. Endsley, "Design and evaluation for situation awareness enhancement," *Proceedings of the Human Factors Society Annual Meeting*, vol. 32, no. 2, pp. 97–101, 1988. [Online]. Available: https://doi.org/10.1177/154193128803200221

[5] ——, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995. [Online]. Available: https://doi.org/10.1518/001872095779049543

[6] B. McGuiness and J. Foy, "A subjective measure of sa: The crew awareness rating scale (cars)," 2000, pp. 286–291.

[7] G. P. Tadda and J. S. Salerno, *Overview of Cyber Situation Awareness*. Springer US, 2010, pp. 15–35.

[8] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, pp. 222–232, Feb 1987.

[9] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, Jan 1997.

[10] C. Onwubiko, "Functional requirements of situational awareness in computer network security," in *2009 IEEE International Conference on Intelligence and Security Informatics*, June 2009, pp. 209–213.

[11] C. Onwubiko, "Understanding cyber situation awareness," *International Journal on Cyber Situational Awareness*, vol. 1, no. 1, pp. 11–30, 2016.

[12] C. D. McDermott and A. V. Petrovski, "Investigation of computational intelligence techniques for intrusion detection in wireless sensor networks," *International Journal of Computer Networks and Communications (IJCNC)*, vol. 9, no. 4, pp. 45–56, 2017.

[13] Y. Zhang, Y. Xiao, M. Chen, J. Zhang, and H. Deng, "A survey of security visualization for computer network logs," *Security and Communication Networks*, vol. 5, no. 4, pp. 404–421, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sec.324

[14] T. Zhang, X. Wang, Z. Li, F. Guo, Y. Ma, and W. Chen, "A survey of network anomaly visualization," *Science China Information Sciences*, vol. 60, no. 12, p. 121101, Apr 2017. [Online]. Available: https://doi.org/10.1007/s11432-016-0428-2

[15] R. E. Etoty and R. F. Erbacher, "A survey of visualization tools assessed for anomaly-based intrusion detection analysis," *DTIC Document, Tech*, Apr 2014.

[16] H. Shiravi, A. Shiravi, and A. A. Ghorbani, "A survey of visualization systems for network security," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 8, pp. 1313–1329, Aug 2012.

[17] L. Axon, S. Creese, M. Goldsmith, and J. R. Nurse, "Reflecting on the use of sonification for network monitoring," in *10th International Conference on Emerging Security Information, Systems and Technologies*, 2016, pp. 254–261.

[18] T. Hildebrandt and S. Rinderle-Ma, "Server sounds and network noises," in *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Oct 2015, pp. 45–50.

[19] L. A. J. R. C. N. M. G. S. Creese, "A formalised approach to designing sonification systems for network-security monitoring," *International Journal On Advances in Security*, 2017.

[20] L. M. Axon, B. Alahmadi, J. R. C. Nurse, M. Goldsmith, and S. Creese, "Sonification in security operations centres: What do security practitioners think?" in *Workshop on Usable Security (USEC) at the Network and Distributed System Security (NDSS) Symposium*. Internet Society, 2018.

[21] A. Rajalakshmi and H. Shahnasser, "Internet of things using node-red and alexa," in *2017 17th International Symposium on Communications and Information Technologies (ISCIT)*, Sep. 2017, pp. 1–4.

[22] C. Z. Yue and S. Ping, "Voice activated smart home design and implementation," in *2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST)*, April 2017, pp. 489–492.

[23] V. Kpuska and G. Bohouta, "Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan 2018, pp. 99–103.

[24] A. Reis, D. Paulino, H. Paredes, I. Barroso, M. J. Monteiro, V. Rodrigues, and J. Barroso, "Using intelligent personal assistants to assist the elderlies an evaluation of amazon alexa, google assistant, microsoft cortana, and apple siri," in *2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*, June 2018, pp. 1–5.

[25] Y. Gao, Z. Pan, H. Wang, and G. Chen, "Alexa, my love: Analyzing reviews of amazon echo," in *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct 2018, pp. 372–380.