Exploring the use of conversational agents to improve cyber situational awareness in the Internet of Things (IoT).

MCDERMOTT, C.D.

2020

Exploring the use of Conversational Agents to improve Cyber Situational Awareness in the Internet of Things (IoT)

Christopher D. McDermott

A thesis submitted in partial fulfilment for the degree
of Doctor of Philosophy
at the School of Computing
Robert Gordon University
Aberdeen, Scotland

October 2020

Supervisor Dr John P. Isaacs and Dr Andrei V. Petrovski

Dedicated to Dominykas

your time with us was short, but thank you for bringing us so much love and joy

xxx

Abstract

The Internet of Things (IoT) is an emerging paradigm which aims to extend the power of the Internet beyond computers and smartphones to a vast and growing range of devices "things", processes and environments. The result is an interconnected world where humans and devices interact with each other, establishing a smart environment for the continuous exchange of information and services. Billions of everyday devices such as home appliances, surveillance cameras, wearables and doorbells, enriched with computational and networking capabilities, have already been connected to the Internet. However, as the IoT has grown, the demand for low-cost, easy-to-deploy devices has also increased, leading to the production of millions of insecure Internet-connected smart devices. Many of these devices can be easily exploited and leveraged to perform large-scale attacks on the Internet, such as the recently witnessed botnet attacks. Since these attacks often target consumer-level products, which commonly lack a screen or user interface, it can be difficult for users to identify signs of infection and be aware of devices that have been compromised.

This thesis presents four studies which collectively explored how user awareness of threats in consumer IoT networks could be improved. Maintaining situational awareness of what is happening within a home network is challenging, not least because malicious activity often occurs in devices which are not easily monitored. This thesis evaluated the effectiveness of conversational agents to improve Cyber Situational Awareness. In doing so, it presented the first study to investigate their ability to help users improve their perception of smart device activity, comprehend this in the context of their home environment, and project this knowledge to determine if a threat had occurred or may occur in the future. The research demonstrated how a BLSTM-RNN with word embedding could be used to extract semantic meaning from packets to perform deep packet inspection and detect IoT botnet activity. Specifically, how the models use of contextual information from both the past and future enabled better predictions to be made about the current state (packet) due to the sequential nature of the network traffic. In addition, a cross-sectional study examined users' awareness and

perception of threats and found that although users value security and privacy they found it difficult to identify threats and infected devices. Finally, novel cross-sectional and longitudinal studies evaluated the use of conversational agents and demonstrated them to be an effective and efficient method of improving Cyber Situational Awareness. In particular, this was shown to be true when using a multi-modal approach and combining aural, verbal and visual modalities.

Acknowledgements

First and foremost, I thank God for giving me the opportunity and strength to complete my doctoral studies, even in the midst of a global crisis and pandemic.

I would also like to thank my supervisory team, Dr John Isaacs and Dr Andrei Petrovski. Thank you for encouraging me to keep going and for not growing tired of my endless questions.

Thank you to Dr Natalie Coull, Dr Cyril Onwubiko and Dr Hatem Ahriz for agreeing to examine my thesis. I look forward to hearing your comments and suggestions.

Most importantly, I would like to thank my family: Wife Emma, and the best children in the world, Jacob, Abigail and Kelvin. Thank you for sacrificing your husband and dad for nearly five years. In the busyness of family life, thank you for not complaining when I was too busy to fix a bike, set up a laptop and build the climbing frame... I promise to spend the next six months completing the list of jobs you have kindly compiled for me. Thank you also to my Mum for making me who I am today, and for agreeing to proof read this thesis.

Finally, I would like to offer some encouragement to anyone who may read this thesis and be contemplating part-time academic study. I share a quote I read in another thesis that encouraged me to keep going and never give up: “More men fail through lack of perseverance than through lack of ability”. Accept there will be bumps along the way, choose to keep going and enjoy the journey.

Declaration

I confirm that the work contained in this PhD project report has been composed solely by myself and has not been accepted in any previous application for a degree. All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.

Signed:  Date: October 2020

Christopher D. McDermott

Contents

List of Tables

List of Figures

List of Algorithms

Listings

Chapter 1

Introduction

This thesis investigates the use of conversational agents for improving Cyber Situational Awareness. In general, situational awareness is defined as *"the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"* [1]. Cyber Situational Awareness is therefore considered the application of situation awareness to the Cyber domain [2]. Previous research has suggested a lack of technical knowledge and ability to explore network communication, results in little or no awareness of security issues in consumer home environments [3]. In this research, the aim was to explore the potential of using conversational agents to help users improve their perception of smart device activity, comprehend this in the context of their home environment, and project this knowledge to determine if a threat has occurred or may occur in the future.

In this first chapter, a background to the research area is presented. Research aims are discussed, a problem statement is defined and the central question addressed by this research is presented. An explanation is also provided regarding how the central question was broken down into sub research questions, each of which is addressed separately in the four main chapters of this thesis (Chapters 4-7). Next, motivation for investigating the research questions is provided, and the contributions this thesis makes to the field of study are described. Finally, the remaining chapters in the thesis are outlined and details of the published literature produced from this thesis are provided.

1.1 Background

The Internet of Things (IoT) has quickly transitioned from a promising future paradigm to a pervasive everyday reality. Billions of smart devices are now being connected to the Internet creating an extensive network of connected *things*, capable of sensing the surrounding environment and interacting with other devices to aid real-time monitoring and decision making [4].

The IoT has now permeated into many areas of everyday life. Three areas of particular growth are in health, industrial applications and smart cities [5, 6]. Central to the future of smart cities is the smart home, where an uptake of low cost and easy to deploy IoT devices, has already been witnessed. This flourishing smart home IoT market is fuelled largely by the promise of convenience, greater interconnectivity and automation of everyday tasks [7]. As a result, smart devices such as *TP-Link's* IP cameras, *Ring's* doorbell and *Philips Hue's* light bulbs, all capable of being switched on using a *conversational agent* such as an *Amazon echo*, are increasingly becoming commonplace in the home. While smart interconnected devices clearly have many benefits, concerns still exist around the security and privacy of such devices, and data derived therein [8]. Many of these concerns arise as a result of device manufacturers excluding security and privacy mechanisms from their products, following market pressure to produce low-cost plug and play smart devices [9]. 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 [10]. 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 [11]. 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 [10]. As previously mentioned, research has suggested a lack of technical knowledge and ability to explore network communication, results in little or no awareness of security issues in consumer home environments [3]. The research in this thesis explores if the rise in popularity of digital assistants and conversational agents, such as the Amazon Echo, could be used to improve Cyber Situational Awareness within consumer IoT networks (smart home).

1.2 Research Aims

1.2.1 Research Question

The aim of this research is to investigate the potential of using conversational agents to improve Cyber Situational Awareness. Specifically, to explore how well participants could use agents to assimilate information about events in their environment (*Perception*), synthesise this into a meaningful understanding of the situation (*Comprehension*) and use the knowledge to identify threats in a home network (*Projection*). The approach adopted by this thesis to achieve this aim is to split the research project into smaller studies, each addressed by a chapter in the thesis. Consideration is initially given to the effectiveness of current detection methods to detect threats before a new method of detection based on deep learning, is proposed. The research next examines user awareness and perception of threats and their ability to detect them within a network. From this, the problem domain can be clearly defined, and a lack of threat awareness confirmed. Next, the viability of conversational agents to achieve the central aim of this thesis is tested. Finally, the utility of conversational agents is tested over an extended period of time, to answer the central research question, which is defined as:

> *Can Situational Awareness of threats in the Internet of Things be improved using Conversational Agents ?*

The central research question is broken down into four distinct research areas. A sub research question is defined for each area and is addressed individually in Chapters 4-7

> **SQ1:** *Can current security methods detect the presence of threats within consumer IoT networks ?* This is addressed in the study of current methods of threat detection in Chapter 4, which focuses on Botnets as being a particular threat facing consumer IoT networks in Chapter 2.
>
> **SQ2:** *Can users visually detect the presence of threats within consumer IoT networks ?* This is addressed in the study of threat perception and awareness in Chapter 5.
>
> **SQ3:** *Are conversational agents a viable method for making users aware of threats in consumer IoT networks?* This is addressed in the cross-sectional study of agent viability in Chapter 6.
>
> **SQ4:** *Are conversational agents effective in making users situationally aware of threats in consumer IoT networks ?* This is addressed in the longitudinal study of agent utility in Chapter 7.

1.2.2 Motivation

This research is motivated by the need to address the growing issue of threats facing consumer IoT networks (smart home). As previously mentioned, research has suggested a lack of technical knowledge and ability to explore network communication, results in little or no awareness of security issues in consumer home environments [3]. If true, the famous quote by Thomas Gilovich, a prominent researcher in social and cognitive psychology, *"People are often unaware of their own unawareness"* [12] can be considered particularly relevant to the context being investigated and a motivating factor to engage in the research.

Personal Author Motivation

As a Computer Scientist I am obviously a strong advocate for technology. I love exploring new ways for technology to make a positive impact in society. I am, however, ever aware of the dangers technology brings. I want to see technology that serves humans not the other way around. Author Cal Newport sums it up best when he says "Technology is neither intrinsically bad nor good. The key is using it to support your goals and values, rather than letting it use you" [13]. As a researcher in human-centred security I am interested in the impact of computing in people's everyday lives, in particular in relation to their security and privacy. My research interest extends to understanding how people perceive their personal information is collected and used; and how quickly they are willing to sacrifice privacy for convenience. More broadly, I am also challenged to find solutions to address the growing trend of technology addiction.

My motivation for undertaking this thesis is to explore one possible way technology can positively impact society, that is, how computing in the new voice era could be used to improve people's perception and awareness of their device activity, to better protect them from risks and threats that exist now and in the future.

1.3 Contribution to the Field

In this thesis, a conceptual Conversational Cyber Situational Awareness Framework is proposed. The main contributions of this thesis are as follows:

1. **Production of a labelled dataset incorporating IoT botnet traffic, and attack vectors (Chapter 4)**. At the time of undertaking this research a lack of IoT botnet datasets was evident, and was therefore a major factor in the decision to create a sandboxed environment and botnet architecture. The generated *mirai* botnet dataset will provide a much needed resource for future researchers in this area, allowing for better understanding of IoT botnets, and the development of new detection methods. The dataset has already been made public and been used for comparative studies [14, 15].

2. **Development of a deep learning method to detect IoT botnet activity (Chapter 4)**. Botnet detection is a research area which has previously received a lot of attention. However, the focus has largely been on traditional networks and not within consumer home environments. The growth of the IoT has seen a rapid proliferation of insecure connected devices across the internet. The huge number of connected devices, coupled with their inherent security issues, has resulted in a surge of powerful distributed denial of service (DDoS) attacks [16], many often now leveraging consumer level products [17, 18, 10]. The deep learning method proposed in this research has many implications for research and industry. It contributes to developing knowledge relating to IoT botnet detection, and provides evidence that deep learning approaches can be successfully applied to this research area. The results of the research have already been used as a baseline for future comparative research and cited extensively[1].

3. **Evaluation of User Awareness and Perception of threats within the IoT (Chapter 5)**. Many studies exist relating to human-centered security and the perception of risk [19]. Understanding how users perceive risk is an important consideration when attempting to evaluate and promote better situational awareness of risks relating to security and privacy. The results from this study contribute to the developing knowledge relating to risk perception and awareness. The contribution has significance since it was clearly demonstrated that users value security and privacy but found identifying threats difficult. The research also demonstrated that a lack of network communication can result in little or no awareness of security issues; however, if presented with data, awareness could be improved.

[1] https://tinyurl.com/rfayhz6

4. **Evaluation of Conversational Agents to improve Cyber Situational Awareness (Chapter 6 and 7).** In recent years conversational agents have experienced a significant rise in popularity, and 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 [20]. Research in this area is growing, and producing some very promising applications of conversational agents. However, this research provides a novel contribution to the developing body of knowledge, since it is the first study to explore the application of aural and verbal analysis, using conversational agents, to the problem of IoT botnet detection.

1.4 Thesis Structure

This thesis consists of eight chapters, the first being this introduction. The remaining chapters are organised as follows:

Chapter 2 Background and Related Work: provides an overview of security within the IoT, exploring in particular the issue of botnets. It briefly looks into the current methods of detecting botnets and their applicability to consumer IoT networks. The chapter explores situational awareness and its application to the cyber domain. Finally, it investigates the use of conversational agents within the IoT.

Chapter 3 Methodology: introduces the research methods used throughout this thesis to answer the central research question. The chapter starts by introducing the philosophy adopted for this research, and explaining the mixed-method approach which was selected. In addition, it also provides justification for the techniques used and explains how they are repeated in several chapters to ensure a level of consistency between studies.

Chapter 4 Botnet detection in Consumer IoT networks: explores a common threat used to leverage insecure smart devices and perform large scale DDoS attacks on the Internet. The taxonomy of an IoT botnet is explored to better understand how infection and spread can occur in smart devices and networks. In addition, the chapter explores the ability of a current detection method to effectively detect botnet activity, before finally proposing a novel application of deep learning for better detection of botnets found within the IoT.

Chapter 5 Situational Awareness of Threats in Consumer IoT networks: examines the awareness and perception of threats within consumer IoT networks. In addition, it analyses user requirements from IoT devices, and the importance placed

upon security and privacy. The chapter also assesses user ability to detect threats within a network, and explores if there is an association between accuracy of detection, and their technical knowledge or age.

Chapter 6 Cross-sectional study to test the viability of Conversational Agents to improve Cyber Situational Awareness: examines the use of conversational agents for improving Cyber Situational Awareness. The chapter presents a cross-sectional viability study which assesses the ability of users to detect threats within a consumer IoT network. A method for assessing situational awareness based on Mica Endsley's SA model is presented, and the results of the study are discussed.

Chapter 7 Longitudinal study to assess the utility of Conversational Agents to improve Cyber Situational Awareness: presents the final study in this thesis, exploring the use of conversational agents for improving Cyber Situational Awareness. Previously, the cross-sectional study collected data from a large population of users at a single point in time. In this chapter, data was collected from a smaller sample of users over an extended period lasting twenty-one days. Mica Endsley's SA model was again used to assess how participants perceive device activity, comprehend this in the context of their environment, and use the knowledge to determine if a threat exists. The results of the longitudinal study are presented and discussed.

Chapter 8 Conclusion: reviews the material presented in the previous chapters. A summary of the findings is presented and the implications of the results discussed. The chapter also discusses the limitations of the research, and provides suggestions for how the research could be extended and taken further.

The relationship between the four main studies of this thesis is presented in Figure 1.1. Each study contributes to the overall narrative, with results informing subsequent studies, and contributing to the central research question.

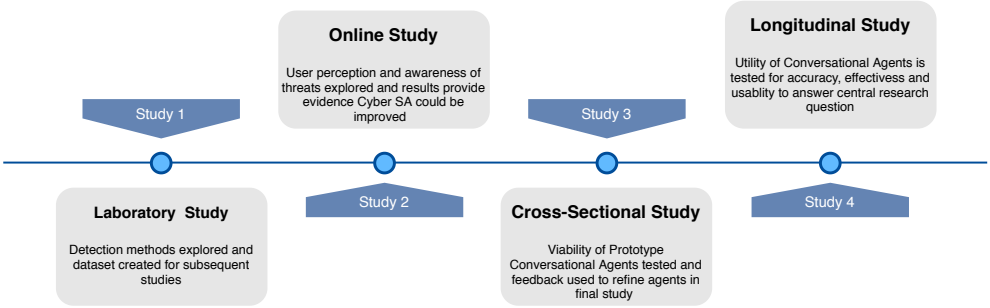


Figure 1.1: Four main studies Chapters 4-7

1.5 Publications

1. McDermott CD, Majdani F, Petrovski AV. Botnet Detection in the Internet of Things using Deep Learning Approaches. In: 2018 International Joint Conference on Neural Networks (IJCNN); 2018. p. 1–8.

2. McDermott CD, Petrovski AV, Majdani F. Towards Situational Awareness of Botnet Activity in the Internet of Things. In: 2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA); 2018. p. 1–8. (Best paper award)

   These papers contributed to the development and application of an algorithm designed in **Chapter 4**, which was used to detect anomalous traffic utilised by the mirai malware. The developed model used a novel application of Deep Bidirectional Long Short Term Memory based Recurrent Neural Network (BLSTM-RNN), in conjunction with Word Embedding, to convert string data found in captured packets, into a format usable by the BLSTM-RNN. In doing so, a solution is presented to address the problem of detecting threats; and making consumers situationally aware if a device is infected and being leveraged as part of an IoT botnet.

3. McDermott CD, Isaacs JP, Petrovski AV. Evaluating Awareness and Perception of Botnet Activity within Consumer Internet-of-Things (IoT) Networks. Informatics. 2019;6(1).

   This paper presented the cross-sectional study in **Chapter 5**, which evaluated how users value and perceive security and privacy in IoT smart devices. It analysed user requirements from IoT devices, and the importance placed upon security and privacy. An experimental setup was used to assess user ability to detect threats, in the context of their technical knowledge and experience. It clearly demonstrated that without any clear signs when an IoT device was infected, it was very difficult for consumers to detect and be situationally aware of threats exploiting home networks. It also demonstrated that situational awareness of threats could, however, be improved if the data was presented to users in an easy to understand manner.

4. McDermott CD, Jeannelle B, Isaacs JP. Towards a Conversational Agent for Threat Detection in the Internet of Things. In: 2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA); 2019. p. 1–8.

5. Improving Awareness of Threats in the Internet of Things using Conversational Agents (under review)

   These papers describe the development of conversational agents for detecting anomalous traffic in consumer IoT networks, presented in **Chapter 6**. The agents accepted inputs in the form of user speech from Amazon Alexa enabled devices and text conversations from a chatbot application. In doing so, the papers presented a solution to the problem of making consumers situationally aware when their IoT devices are infected, and anomalous traffic has been detected.

6. Intrusion Detection using Mulimodal Analysis in the Internet of Things (under review)

   This paper presents the results of a longitudinal study in **Chapter 7** where the utility of conversational agents was assessed. The study was mapped to Mica Endsley's Situational Awareness model and was used to assess how participants perceive device activity, comprehend this in the context of their environment, and use the knowledge to determine if a threat exists. In addition, the usability of the agents was evaluated for effectiveness, efficiency and satisfaction.

Chapter 2

Background and Related Work

In this chapter, existing research is reviewed for its significance and relevance to the research presented in this thesis. First, the Internet of Things (IoT) is defined and the importance of the paradigm explained. Next, security concerns within the IoT are highlighted, with a particular focus on IoT malware and botnet activity. A critical review of the current methods of detecting DDoS attacks and botnet activity is also presented. Situational awareness is then defined and contextualised for the Cyber domain. Finally, the growth of voice computing and the adoption of conversational agents is discussed, exploring current applications of the technologies, and identifying gaps in the literature relating to its application to Cyber Situational Awareness.

2.1 Internet of Things (IoT): Domains and Security Considerations

The Internet of Things (IoT) is a novel paradigm that has the potential to revolutionise large sections of everyday life. At its core, the aim of the IoT is to connect previously unconnected devices to the Internet [5], thus creating smart devices capable of collecting, storing and sharing data, without requiring human interaction [4, 21]. The term *Internet of Things* was first coined by Kevin Ashton in 1999 in a presentation made to multinational consumer goods company Proctor and Gamble [22]. In his presentation, Ashton proposed linking RFID in P&G's supply chain to the internet. His vision was clear:

> "*If we had computers that knew everything there was to know about things - using data they gathered without any help from us - we would be able to track and count everything, and greatly reduce waste, loss and cost. We would*

know when things needed replacing, repairing or recalling, and whether they were fresh or past their best."

This initial view of the IoT was influenced heavily by the focus of the Auto-ID Labs network[1], a research group in the field of networked radio-frequency identification (RFID) and emerging sensing technologies. Their aim was to develop an Electronic Product Code (EPC), giving physical objects a globally unique identifier, so that when coupled with RFID technology, an objects visibility (status, current location) can be tracked and monitored at all times [5]. While this view certainly describes an important part of the IoT, it does not reflect the full vision, since it limits *things* within the paradigm to RFID tags. Indeed, alternative definitions of the IoT recognise that the term *IoT* implies a much wider vision than just object identification. Technological advancements in electronics and computing have led to an exponential increase in internet connected *things*, and a widening of the application domains covered by the IoT [21]. Mosenia et al. suggest the scope of IoT applications includes: Smart Vehicles, Smart buildings, Health monitoring, Energy management, Construction management, Environment monitoring, Production and assembly, and Food supply chains [21]. Gubbi et al. categorise the applications into four application domains: (1) Personal and Home; (2) Enterprise; (3) Utilities; and (4) Mobile [23]. Similarly, Atzori et al. suggest the potential applications of the IoT are numerous and propose categorising them into four similiar application domains: (1) Transportation and logistics; (2) Healthcare; (3) Smart environment; and (4) Personal and social [5].

Although, it could be argued that categorising the scope of IoT applications is subjective, a number of limitations were identified in the existing literature. First, although Mosenia et al. discussion of IoT applications was extensive, further refinement could produce a more succinct categorisation. Conversely, Gubbi et al. categorisation was too narrow, resulting in a list which did not reflect the breadth of possible applications. Finally, Atzori et al. was found to be the most complete in terms of breadth of coverage and succinctness, however, did not adequately cover utilities and clearly reflect leisure activities. The application domains were, therefore, reorganised and categorised as shown in Figure 2.1.

[1] https://www.autoidlabs.org/

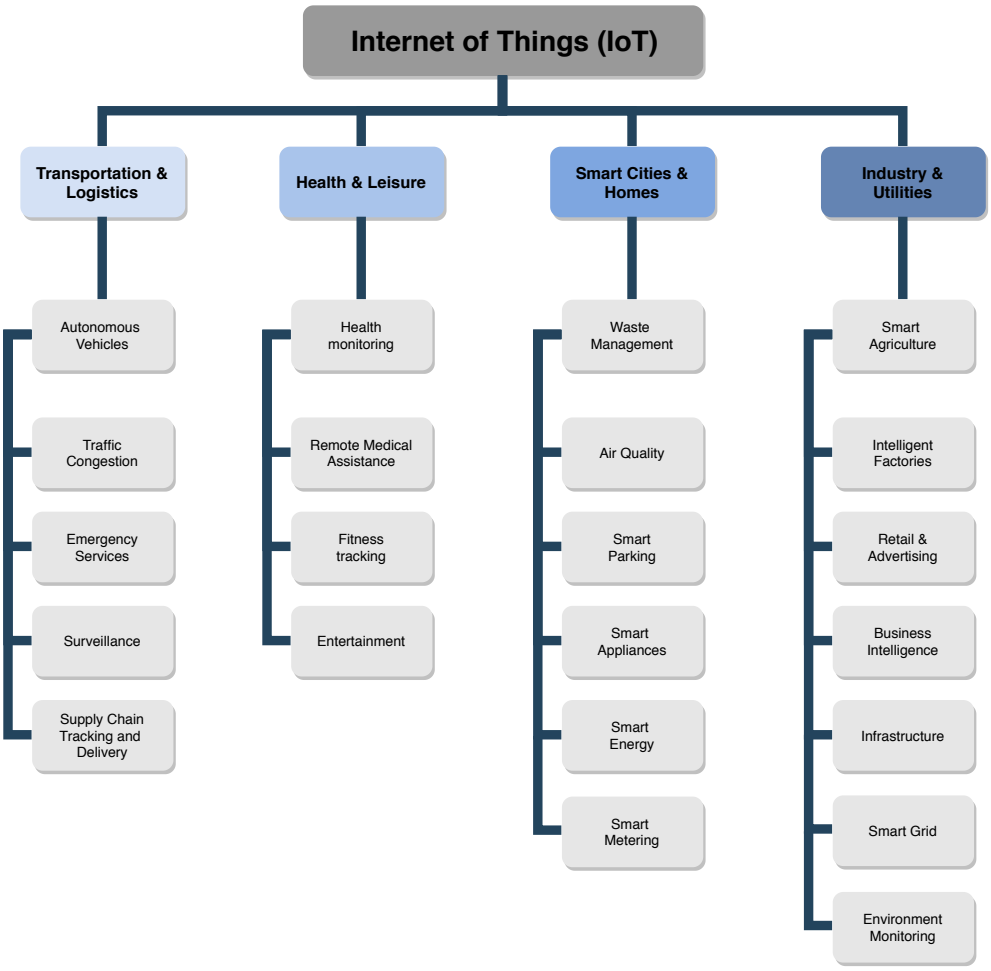


Figure 2.1: IoT Application Domains

Transportation and Logistics. The first category covers any activity of transporting people or goods and services to customers. Traffic congestion and road accidents have long imposed intolerable burdens on drivers [24]. Traditional solutions to such problems have included: increasing road capacity, reduction of demand through road tools, and promoting greater car sharing. However, smart and autonomous vehicles have started to revolutionise traditional transportation, helping to address some the associated issues [21]. The transportation improvements have the potential to greatly benefit Emergency services. With fewer road accidents, and less congestion, demand for services could be reduced and logistical operations of ensuring timely responses to incidents and emergencies optimised. The IoT has the flexibility to provide different levels of vehicle autonomy depending upon the situation. Automated vehicle systems such as *pilot assist* can assist in addressing issues previously discussed. However, fully autonomous (self-driving) vehicles, have the potential to provide even further benefits [25]. For example, supply chain tracking and delivery can be better monitored and

managed using autonomous vehicles, as demonstrated by *Amazon*[1] and their development of autonomous delivery robots. In addition, the use of unmanned aerial vehicles (UAVs), is now being explored as another method of reducing cost and time required to deliver packages [26, 27].

Health and Leisure. The second category covers various automation processes relating to health and leisure which can be improved through expansion of the IoT. For instance, the evolution of health monitoring systems over the past two decades has the potential to change the way health care is currently delivered [28]. Health professionals are able to monitor patients, particularly older adults [29] for conditions such as dementia [30] and Alzheimer's [31]. With an aging population and more people living alone, remote home monitoring made possible by the IoT, will ensure elderly people are able to maintain their independence and quality of life [29]. Remote monitoring however is not limited to the elderly with wearable health monitoring systems designed to track fitness and vital statistics such as heart rate (HR), blood pressure (BP), electrocardiography (ECG), oxygen saturation (SpO2), body temperature and respiratory rate (RR) [28, 32]. The benefits derived from the IoT extend beyond the home, and make remote medical assistance increasingly possible. For example, drones are now used as part of the medical-supply infrastructure to provide help and deliver resources to remote locations that lack adequate roads [33]. In [27] fixed-wing drones were used to deliver blood and life saving medicine in Rwanda. In [34] following a magnitude 7.8 earthquake in Nepal (2015) drones were deployed as part of the humanitarian effort to provide vital and real-time information to rescue teams. The benefits the IoT brings to healthcare are clear, and will likely increase as the IoT grows and expands. Finally, entertainment and social networking are about to be transformed through connectivity to billions of interconnected *objects*. RFIDs can be used to collect information about our social activities and upload real-time updates to social networks, replacing the manual "Checking in" process often used with sites like *Facebook*, in order to let friends know where you have been [5].

Smart Cities and Homes. The third category covers various automation processes relating to smart cities and homes. IBM define a smart city [35] as

> "*the physical infrastructure, the information-technology infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city.*"

Alavi et al. broaden this definition of a smart city to an urban environment which

[1] https://tinyurl.com/amazon-autonomous-vehicles

utilises technologies and digital data to deliver better public services, through more effective use of resources [36]. Predictions suggest that by 2050 66% of the global population will be living in urbans areas, many in mega-cities of over 10 million inhabitants[1]. It is therefore agreed that the establishment of *Smart Cities* is a core requirement to cater for the expected rapid global urbanisation [37]. Expanding modern cities face challenges relating to management, efficiency and quality. The smart city, fuelled by the IoT, is set to provide solutions to many of these urban challenges such as waste management[38], air quality[39], smart parking [40] and energy consumption [36]. A basic building block for smart cities is the *Smart Home* [37]. A Smart home is defined as a building that contains a communications network (gateway) to connect appliances and services, allowing them to be remotely controlled, monitored and accessed [41]. The gateway is commonly accessed through an application running on a tablet, mobile phone or computer; and is used to control heating, lighting, ventilation and security systems within the home. Control of appliances is not limited to a user interface, with voice command systems such as [42] being developed as multi-functional *Smart Home Automation System (SHAS)* to control doorbells, fans, lights and curtains within the home; and security and sprinkler systems in the garden. In addition, systems controlling monitored appliances can also be set to post notifications to social media platforms such as *Twitter* [43]. The result is a home where the efficient control of the building and appliances allows homeowners to remotely monitor the status and environment of their home, and have real-time control of connected *objects* (temperature settings, door locks, security cameras, etc.).

Industry and Utilities. The final category covers IoT applications within industry and utilities. Industrial processes can be added to the IoT to create fully autonomous operations, where groups of devices work together to achieve a process. For example, building automation systems (BAS) can be added to the Internet to connect to existing infrastructure and *Smart Grids* for better device maintenance and energy efficiency [44]. Likewise, in agriculture, key-systems such as irrigation can be integrated into the IoT to enable intelligent control. In addition, the tractability of produce and movement of animals can be tracked using IoT technologies, particularly useful during the outbreak of a contagious disease [45]. Product tracking and logistics is not limited to agriculture, and can also be used to enhance retail. Products equipped with RFIDs can be integrated with smart shelves to allow real time monitoring of stock and detection of shoplifting. Business intelligence and advertising can combine to create new opportunities for targeted adverts. For example, in [46] the authors suggest the advent of autonomous vehicles will move the car from a mechanical machine used to move people

[1] https://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html

between two points, to a dynamic and targeted environment to build novel applications and services. If for example, a mechanism existed to *profile* passengers, since the car is connected to the smart city, which in turn is connected to shops, restaurants and bars, promotion adverts could be displayed inside the car as it passed key locations. Finally, as industry and utilities are added to the IoT within smart cities, environment monitoring will become critical to ensure quality of life within smart homes and cities.

2.1.1 IoT Security

The Internet of Things (IoT) is expected to usher in an era of increased connectivity, with billions of devices expected to be connected to the Internet [47]. Many of the smart devices found within the IoT are aimed at consumers, who value low cost and ease of deployment over security. As a result, these market forces have resulted in IoT manufacturers omitting critical security features, and producing swathes of insecure Internet connected devices, such as IP cameras and Digital Video Recorder (DVR) boxes [11]. Such vulnerabilities were investigated in [48] where three popular IoT devices were tested: *Philips Hue* light bulb, *Belkin WeMo* power switch and *Nest* smoke alarm. The authors undertook extensive analysis of device activity and communications, and demonstrated a clear lack of encryption, appropriate authentication and privacy concerns when using the devices. In [49] vulnerabilities were also identified in IP cameras and a smart home toilet, which used a default Bluetooth password, allowing it to be controlled by anyone with the associated app. These highlighted vulnerabilities and exploits are often derived and epitomised by smart device characteristics such as inherent computational limitations, use of default credentials and insecure protocols. The rapid proliferation of insecure IoT devices and ease by which attackers can locate them using online services, such as Shodan, provides an ever expanding pool of attack resources. By compromising and leveraging multitudes of these vulnerable IoT devices, attackers have the potential to perform large scale attacks such as spamming, phishing and Distributed Denial of Service (DDoS), against resources on the Internet [50]. It is clear, as smart homes increasingly adopt IoT devices, it is vital to develop specific security solutions for the IoT to enable users and organisations to protect their smart devices better [51].

2.1.2 Botnets in the Internet of Things

Some of the most extensive and destructive cyber-attacks deployed on the Internet have been DDoS attacks [52]. Figure 2.2 presents a timeline of prominent malware which specifically targeted IoT devices, leveraging them to perform large scale DDoS attacks. Several of these attacks, including the largest ever to be recorded, occurred in the second half of 2016, fueled in full or part by the IoT. During this time, attacks

of over 100 Gbps significantly increased by 140%, with three attacks reaching over 300 Gbps [16]. Attacks of the same severity and magnitude continued into 2017 and by the fourth quarter of 2017 Verisign also reported that 82% of DDoS attacks now employed a multi-vector attack strategy. This evidence suggested IoT botnets were becoming increasingly more common and sophisticated in their effectiveness and ability to exploit basic security vulnerabilities, and obfuscate their activity [16]. Indeed, the growing trend of IoT malware has continued as demonstrated in Figure 2.3b, which shows an increase in the number of IoT malware samples between 2017-2019. With 20.4 billion devices forecast to be connected to the Internet in 2020, maintaining security and privacy within Smart Homes and Cities continues to be a challenge. Therefore, to better understand this growing challenge, a brief history of malware specifically targeting the IoT is presented. Following De Donno et al. [53, 54], only prominent examples which leverage IoT devices to perform DDoS attacks between 2008-2018 are discussed. Therefore, like the authors, other IoT malware with different goals are also omitted. This literature review complements De Donno's original list, adding additional recent examples found within the literature.

Linux.Hydra. The earliest known malware specifically targeting devices found within the IoT [55]. It was managed by IRC, which was historically a popular method to host botnets, due to the networks simple, and low bandwidth communication methods. The malware targeted routing devices based on MIPS architecture, gaining access through a brute force dictionary attack, or leveraging a D-Link authentication bypass exploit [56].

Psyb0t. Another IRC malware which targeted MIPSel architecture, common on network equipment running Linux-based operating systems such as OpenWRT and DD-WRT [56]. Since router firmware is usually read-only, the malware could only run in RAM, however, proliferate was swift. Access was gained through brute force dictionary attacks or by leveraging a D-Link authentication bypass exploit. Once infected the malware could be used to initiate DDoS attacks or access other services such as MySQL, FTP and SMB.

Chuck Norris. Malware which appeared in 2010 and shared code and functionality with its predecessor (Psyn0t). Like Psyb0t, the binary was IRC-based and targeted network devices running on the MIPSel platform. Its method of encrypting information and proliferation was similar to Psyb0t scanning a list of IP addresses stored in a file on the router, or a hard coded list within the binary [54].

Tsunami. Another malware binary which appeared in 2010 sharing code with its predecessor (Chuck Norris), and also distinguishing features associated with the *Linux*

Kaiten/Tsunami open source DDoS tool [56]. The binary was still IRC-based, highly favoured for its means of C&C communication protocol. In addition to previous attack vectors, the malware included capability to perform *HTTP Layer 7 Flood* attacks where genuine HTTP GET or POST requests are used to perform DDoS attacks [57]. *TCP XMAS* attacks were also included allowing malformed packets, with all flags enabled to be created, and overload a target [58].

Aidra. Developed in 2012 the malware, also known as *LightAidra*, did not appear to follow the same development path as the previous binaries, and targeted a wider range of architectures (*ARM*, *PPC* and *SuperH* [54]. Still IRC-based, delivering the same attacks as previous malware, however was able to modify firewall settings using *iptables* [59]. Its use of a cross-compiled binary, to infect multiple IoT binaries at the same time, was adopted in subsequent malware binaries [60].

Spike. The first notable IoT malware employing a new architecture model. Previous malware specifically targeting the IoT utilised IRC for communicating. An IRC channel was established for infected clients (*bots*) to join, and commands were sent to the channel via an IRC server. Bots would receive the commands, execute the instructions, and return their results to the IRC channel. *Spike* malware, however, used an agent-handler model, where software packages (*handlers*) were setup on a server, and *agents* installed on infected IoT devices. The attacker communicates with handlers to identify available *agents*, and instructions are sent to the agents to perform the required attack [61].

Bashlite. The second prominent example of agent-handler malware targeting the IoT was first detected in 2014. Although classified as using an agent-handler model, *Bashlite* C&C operate similar to IRC channels, to allow operators to interact while connected to the C&C [53, 62]. Interestingly, the C&C IP addresses are hard-coded, making the malware easier to monitor [63]. Available attack vectors are in line with previous malware, however, additional architectures are also targeted, including *SuperH* and *SPARC*. The sourse code was released onto the Internet, leading to many variants being subsequently developed.

LizardStresser. The malware came to prominence shortly after Christmas Day 2014 when it was used to bring down the Sony Playstation and Microsoft Xbox gaming networks [64]. The malware targeted *x86, ARM* and *MIPS* CPU architectures, commonly used on embedded IoT devices. Once infected targets were used to scan the Internet for further targets accepting connections via Telnet, and a brute force attack was performed checking against a list of known default credentials.

Elknot. Also known as *BillGates* the malware gained a lot of traction in China in

2015. The malware targeted similar architectures to previous malware, but added a new DNS Amplification attack, where open DNS resolvers were leveraged to overwhelm a target with an amplified amount of traffic [61].

XOR.DDoS. The second prominent malware targeting IoT devices detected in 2015. The malware was able to exploit the Shellshock vulnerability, although it did not rely upon the vulnerability to gain access [54]. The Shellshock vulnerability was a security bug causing Bash to execute commands from environment varaibles. Attackers were able to remotely issue commands on a server, a process also known as remote code execution. Attack vectors including DNS amplification.

LUABOT. The first IoT malware written in LUA programming language, a lightweight embedded scripting language. Since the language is cross-platform it proved effective for exploiting systems running an embedded version of linux. The binary prepared targets to be centrally controlled by the botnet, and was able to copy device configurations and certificates to be sold for use in cloned devices [65]. Once infected, remote access to a target is blocked through the use of tailor made *iptables*.

Remaiten. Detected in 2016, the malware presented as a fusion of two previous malwares: *tsunami* and *LizardStresser*. The method of proliferation, scanning for available *telnet* connections, was borrowed from *LizardStresser*, while the handling of C&C messages was borrowed from *tsunami*. As such C&C communications used an IRC Channel, but just the IRC protocol [53].

New Aidra. Existing malware continued to be fused together in 2016 to create new, more potent variants. The original *Aidra* root code, was combined with *tsunami* IRC-based approach, *BASHLITE* scanning/injection and *Mirai*'s use of a dictionary attack to create *New Aidra*, also known as *Linux.IRCTelnet*. Released around the same time as *Mirai*, it demonstrated how developers built upon existing malware to create new, and more powerful campaigns.

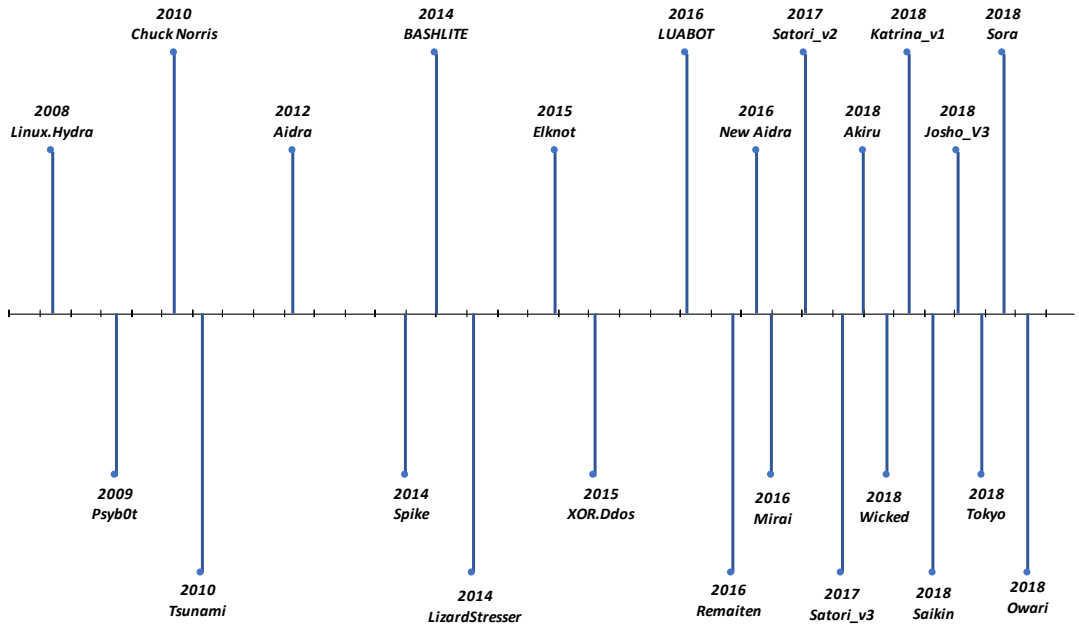


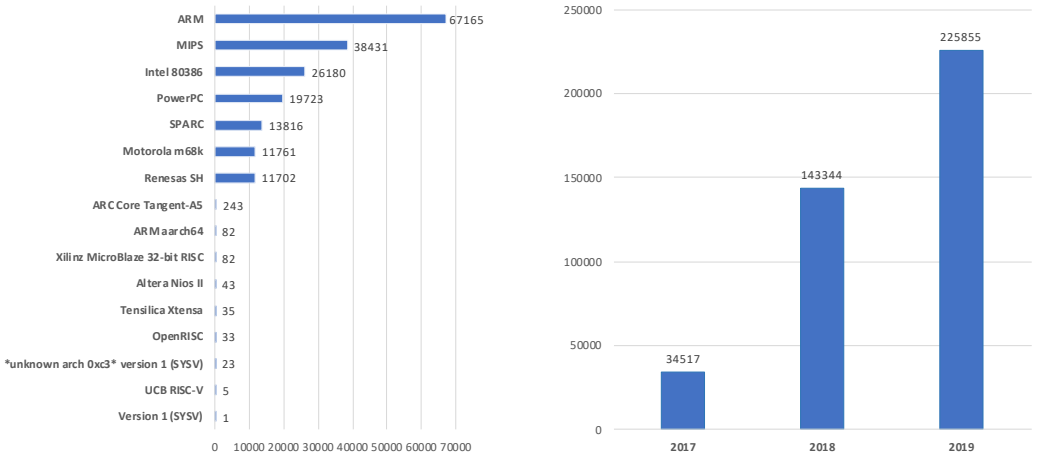
Figure 2.2: IoT DDoS Malware Timeline

Mirai. Arguably the most prominent example of IoT malware in recent times was first detected in 2016 [17, 18, 10]. On 20 September 2016, the *Mirai* botnet was used to perform an unprecedented 620 Gbps DDoS attack on security journalist Brian Krebs website krebsonsecurity.com [66]. Shortly after, it was also responsible for a series of additional DDoS attacks against French web hosting company OVH, and DNS provider DYN. Peaking at over 1.2 Tbps, it was estimated that up to 100,000 infected IoT devices (bots) were involved in the attacks [65]. The severity of the DYN attack was sufficient to cause major disruption on the Internet, and render several high-profile websites such as GitHub, Twitter, Reddit, Netflix inaccessible. This proved to be a watershed moment, defining the future of IoT malware to the present day. Following the release of source code on GitHub [1], *Mirai* has quickly become the framework for malware targeting devices in the Internet of Things [60]. Indeed, the impact of *Mirai* has continued to dominate the landscape into the second half of 2019, with tens of thousands of unique versions of the malware detected on the Internet, an increase of 57% from 2018 [67] (See Figure 2.3b). Although at first glance *Mirai* appears to utilise many features seen in previous malware, the implementation of some (features) appear to be more sophisticated. In [60] the authors identified a number of *Mirai* enhancements such as its use of a random subset of credentials when performing brute force dictionary attacks. Also, since IoT malware resides in RAM, and a simple restart of the device

[1] https://github.com/jgamblin/Mirai-Source-Code

can remove the malware, *Mirai* introduced an new anti-reboot feature. In addition, *Mirai* introduced a new stateless scanning method when looking for new targets. The malware no longer needs to wait for a timeout to execute before moving on to a new IP. Once a target is found, *Mirai* also now determines the device architecture and only sends the corresponding binary, an improvement from previous malware which downloaded all binary variants.

Mirai Mutations (2016-2018). As mentioned previously, following the release of the *Mirai* source code, it quickly became the framework for new malware targeting the IoT [60]. In 2019, Netscout reported detecting 225855 samples of *Mirai* variants [67]. The vast majority of them targeting the same architectures (*ARM, MIPS, Intel, PowerPC* and *SPARC*) as the original malware (See Figure 2.3). In addition, the top five exploits used by the malware continue to be: *Huawei Router HG532*, *Realtek SDK*, *Hadoop YARN Resource Manager*, *D-Link DSL*, and *Linksys E-series*. For brevity, the original list of malware compiled by De Donno et al. [53, 54] is complemented with ten new prominent malware, all variants of *Mirai* (See Table 2.1).



(**a**) Targeted Operating Systems

(**b**) Unique Mirai Samples

Figure 2.3: Mirai Malware Variants

In [68] the authors present *Akiru, Katrina_V1, Sora, Saikin, Owari, Josho_V3* and *Tokyo*, all new mutations of the original *Mirai* malware. Indeed, the proliferation is also evident in [69, 70] where *satori*, *masuta*, and *wicked* are presented. Sharing the original basecode with *Mirai*, many features and characteristics are retained from the original as shown in Table 2.1. For example, the string used to confirm whether the malware is already present on a device, or should be loaded, has only seen minor alterations such as "*MIRAI: applet not found*" to "*Akiru: applet not found*". Although

only a minor difference the change can have an impact on the effectiveness of Intrusion Detection Systems (IDS) signatures[1] to detect *Mirai* such as

alert tcp $EXTERNAL_NET any <> $HOME_NET any (msg:"Possible Mirai infection"; content: "MIRAI: applet not found"; sid: 10003; rev:1;)

Here, the *Snort* rule uses string *MIRAI: applet not found* to trigger an alert, whenever a match is observed. While this rule would be effective for the original malware, the rule would require modification for each variant of *Mirai*. In addition, Table 2.1 also shows that in many cases new variants have been enhanced to scan for additional ports, architecture types (*ARC*), and exploits. Further enhancements also include the ability to take direct control of compromised devices, making other malicious actions possible, including running trojan viruses, redirecting traffic for man-in-the-middle attacks, and delivering other viruses to devices on the network by proxy [68]. Some new variants also target smart signage TVs and wireless presentation systems, such as *LG* Supersign TVs and the *WePresent* WiPG-1000 wireless presentation system. [71].

Table 2.1: Mutated Mirai Malware

	Infection	**Exploit**	**Ports**	**Arch**	**Date**	**Note**
Mirai	MIRAI: applet not found	CCTV-DVR. Netis Router. RealTek SDK.	23 2323	MIPS ARM PPC SPARC	09-2016	Original Malware used to attack krebsonsecurity.com
Satori_V2	nexus was here	CCTV-DVR. Netis Router. RealTek SDK.	81 53413 52869	ARC RCE	08-2017	Additionally suspected to scan for D-Link DSL-2750B and XionMai uc-httpd 1.0.0 devices
Satori_V3	nexus was here	CCTV-DVR. Netis Router. RealTek SDK.	81 53413 52869	SuperH	12-2017	Possibly also exploits UPnP vulnerability in Huawei routers
Akiru	Akiru: applet not found	CCTV-DVR. Netis Router. RealTek SDK.	81 53413 52869	ARC RCE	01-2018	One of three new variants to target ARC (Argonaut RISC Core) and RCE (Motorola RCE) architectures
Wicked	WICKED: applet not found	CCTV-DVR. Netgear SOHO. Huawei HG532.	8080 8443 80 81	-	05-2018	Suspected to be same creator as OWARI due to his handle name

[1] https://www.snort.org/rule_docs/1-40519

Katrina_V1	Katrina: applet not found	Netis Router. RealTek SDK. Huawei HG532.	53413 52869 37215	-	06/2018	No additional information
Saikin	Saikin: applet not found	-	-	ARC RCE	06/2018	Contains eighty passwords, with only four previously used in the original malware
Josho_V3	daddyl33t: applet not found	-	-	-	06-2018	daddyl33t is reference to the developer who created QBot botnet
Tokyo	MIRAI: applet not found	Netis Router. RealTek SDK. Huawei HG532.	53413 52869 37215	-	06-2018	Only variant to use the default Mirai check-string
Sora	Sora: applet not found	Netis Router. RealTek SDK. Huawei HG532.	53413 52869 37215	-	07-2018	Project abandoned but later developed by new hacker using Aboriginal Linux
Owari	Owari: applet not found	Netis Router. RealTek SDK. Huawei HG532.	53413 52869 37215	-	07-2018	Suspected to be same creator as WICKED due to his handle name

- denotes no change from original Mirai malware

Other Notable IoT Malware. A number of other notable IoT malware exist, however were not included in Figure 2.2 and Table 2.1 since they either only share a portion of code from *Mirai* or do not currently leverage devices to perform DDoS attacks. *JenX* shares characteristics with previous malware including exploit vectors seen in *Satori* and *Masuta*. Interestingly, the malware used central C&C servers, hosted on a site providing services for Grand Theft Auto, to perform scanning and exploitation. This is a different approach from most other IoT malware which use distributed scanning and exploiting, where each target once infected performs its own scan to find new targets to infect [72]. *Brickerbot* was first discovered in 2017 and leveraged SSH default credentials of vulnerable IoT devices to perform a permanent denial-of-service (PDoS) attack. The malware attempted to gain access by brute force using default telnet passwords, execute commands using *busybox* to corrupt MMC and MTD storage, before deleting all files, and disconnecting the device from the Internet [65]. Interestingly, subsequent versions

of malware used the *Tor* network to conceal its location and IP address, and did not rely on the presence of busybox on the target device. *Hajime* is presented in [73] which appeared at the same time as *Mirai*, but is considered to include more sophistication. In addition to scanning TCP port 23, the malware is also able to attack port 5358 which provides a way to implement web services on resource constrained embedded devices [65]. In addition, the malware supports Universal Plug and Play (UPnP) an Internet Gateway Device (IGD) protocol supported by many NAT enabled routers. The malware's use of fully distributed communications and UDP port 1457 enables it to make use of the BitTorrent protocol for peer discovery.

The rise in IoT based DDoS attacks, witnessed in recent years, will likely continue until IoT manufacturers accept responsibility and incorporate security mechanisms into their devices. Until such a time, the IoT has the potential to become the new playground for future cyber attacks and therefore presents a number of challenges.

2.1.3 Intrusion Detection Methods

The previous section highlighted prominent examples of malware which specifically targeted devices found within the IoT. It also demonstrated that with 20.4 billion devices forecast to be connected to the Internet in 2020 maintaining security and privacy within Smart Homes and Cities continues to be a challenge [67]. This section explores methods for detecting malware within IoT environments. Specifically, the use of *Intrusion Detection Systems* (IDS) as a form of passive network monitoring, in which traffic is examined at a packet level and results of the analysis are logged. In doing so, they can be an effective countermeasure against botnet activity by observing and identifying active attacks and vulnerabilities in network traffic [74].

Network security monitoring has long been a topic of research, ever since Anderson [75] published his seminal work where he demonstrated how audit logs could be analysed to identify anomalies. Denning [76] continued this work producing a framework for a general purpose intrusion detection expert system. Again, the focus was on identifying security violations by monitoring a system's audit records. In this respect, the author regarded the model as a rule-based pattern matching system, since when an audit record was generated, it was matched against a profile for normal and abnormal behaviour. Lunt et al. [77] proposed improvements to Denning's approach by suggesting the use of *prioi* rules of "socially unacceptable" behaviour where a legitimate user who may abuse their privileges, and engage in activity outwith their normal behaviour. Their proposed system would identify this activity as separate anomalies to genuine intrusion attempts by an attacker. Heberlein et al. [78] extended the work of Lunt to cover intrusions at the network level, focusing on security related issues in a single broadcast

LAN. The proposed Network Security Monitor (NSM) measured network utilisation and host-to-host activity and used probabilistic, rule-based and mixed approaches to detect anomalous behaviour. They further extended their work by proposing a Distributed Intrusion Detection System (DIDS) [79], which combined host based intrusion detection with network traffic monitoring. They also suggested that bench-marking mechanisms should be developed in order to test the effectiveness of intrusion detection systems. DARPA duly complied [80] and developed an intrusion detection test bed which generated normal and attack traffic. Six research groups participated in a blind evaluation of intrusion detection systems to evaluate their effectiveness. The idea of a Distributed Intrusion Detection System was developed further in [81], where a multi-sensor data fusion approach was proposed. Here, data from multiple sources is combined to make inferences about events, activities and situations. In doing so, Bass hypothesised that the fusion of data from a myriad of distributed network sensors could provide a framework for building next-generation IDS and effectively achieve "Cyberspace Situational Awareness". The concept of Cyber Situational Awareness will be discussed in Section 2.2.2. Since these seminal pieces of work a number of approaches have been proposed. Following Zarpelao et al. [82], intrusion detection techniques presented in this thesis are categorised into three main approaches: signature-based, anomaly-based, specification-based. Hybrid approaches also exist, but generally have a dominant method, therefore, these are categorised into one of the three main approaches. Relevant literature relating to the use of each approach is discussed and interesting themes and use cases are highlighted, such as *whitelisting, behavioural anlaysis, Software Define Networking* and *Blockchain* technology for detecting botnet activity and DDoS attacks.

Signature-based Detection. Commercial intrusion detection systems are predominantly signature-based, detecting attacks by comparing a known attack pattern (signature) to incoming attacks [83]. As such they can be an effective tool for detecting known threats, but require frequent rule-base and signature updates. Kambourakis et al. [65] suggest strong signatures can be an effective method of detecting malware such as *Mirai*, *Hajime* or *Bashlite* presented in Section 2.1.2. For example, in the case of *Mirai* they suggest rules could be written to monitor ports 23, 2323, and 22 for repeated authorisation attempts. Alternatively, patterns of activity could be monitored such as , (1) establishment of a TCP connection, (2) transmission of a sequence of packets of predictable size, (3) termination of the connection. Patterns matched to a pre-defined signature would then raise an alert.

Whitelisting has also been used as a mechanism to identify specific devices and explicitly allow access or privileges to perform functions [84]. Gopal et al. [85] propose a system for detecting and mitigating the spread of IoT malware. The system utilises

application whitelisting to allow only trusted applications to be executed. In the context of IoT malware, the application is the binary used by IoT malware to infect and propagate within a network. The system is built using a modular architecture, with the first module (*Profiling*) responsible for scanning the router or IoT node for binaries which require to be executed. Once found, a hash value is calculated for each binary and stored in a database. The second module (*Enforcement*) uses the hashed signatures to enforce the whitelist of trusted binaries. The module computes the hash of the binary just prior to execution, and compares the value to the hashed signature in the database. If the values match, the binary is allowed to execute, otherwise it is blocked. Testing and results were limited, but appeared to demonstrate some promise when using whitelisting as a method of IoT malware detection and prevention. Indeed, the same approach is taken in [86] where *Heimdall*, a whitelist-based intrusion detection approach, is used to detect common IoT attacks, such as DDoS and passive eavesdropping. Similar to the approach used by Gopal, a profile is created for each monitored IoT device, however, here the whitelist is composed of legitimate contacts (other IoT devices) a device can make in order to perform its functions. In doing so, the system can prevent incoming attack traffic from logging into a device during the infection stage of malware propagation. In addition, the whitelist also governs outgoing traffic, preventing a device from communicating with illegitimate destinations (botnet servers such as C&C). The whitelist is initially generated by monitoring DNS traffic for each device, and storing hostname/IP mappings to a device profile, constituting trusted destinations. If monitored traffic matches the profile of a device, the traffic is allowed, otherwise it is blocked. The system was evaluated for its effectiveness to protect five popular IoT devices including a Smart Lock and Light bulb, however the evaluation only focused on validating the overhead of running the system on an edge router. There was no clear indication whether attempted connections made by IoT malware were successful or not. In addition, during the generation of the initial whitelist, consideration was not given to an event where an infected IoT device was already present on the network. This could potentially result in malicious traffic being identified as trusted, and included in the whitelist.

Gu et al. [87] present *BotHunter* a perimeter detection system which focuses on detecting malware infections associated with botnets. The IDS is built on top of *Snort*[1] and monitors ingress/egress traffic matching it against an extensive set of malware-specific signatures. In addition, two custom plugins were developed to complement Snort's signature engine and provide inbound and outbound scan detection warnings that are weighted for sensitivity toward malware-specific scanning patterns. Ten different IRC-based botnet variants were used to test the system (*Agobot, Gaobot, Phatbot, SDBot,*

[1] https://www.snort.org/

RBot, UrBot, UrXBot, and *GTbot*. Extensive testing in virtual and live environments was undertaken, demonstrating that the system is capable of accurately flagging both well-studied and emergent bots. However, some limitations are evident. Firstly, the premise of the system is to detect botnet activity by monitoring C&C communications, however if future botnets encrypted their communications the system would need to be adapted. It could also be possible to evade detection by using different paths for inbound and outbound traffic (rerouting) which would pose another challenge to the system. Monitoring C&C communications channels to detect botnet activity was also used in [84], where the authors monitored traffic for unusual or suspicious IRC nicknames, matching these against pre-defined signatures.

In [88] Hadi et al. the authors propose a botnet detection system (*BoDMitM*) which uses policies defined for IoT devices, to determine the level of network access granted to each device. The policies are defined using Manufacturer Usage Descriptions (MUD)[1] which provide device visibility and allow the system to identify each IoT device type and define the appropriate behaviours for the device. The system is configured to run on *Openwrt*[2], a popular firmware for SOHO routers, and uses *Snort* as the detection engine. When a new IoT device is added to the network, the device uses its MUD to inform the network what type of device it is, and what network access it requires to perform its function. The system monitors the network traffic and attempts to match the MUD of new devices to a MUD policy. If a match is found, access is granted according to the specified policy, otherwise a violation occurs, and the traffic is forwarded to the IDS (Snort) for further analysis. The authors reported 100% efficiency in detecting attacks, however, did not provide details of the testing procedure, or how the IDS handled the malicious traffic. In addition, the system is built on the premise that malware enters the network when new infected devices are added, however, in reality, devices already in the network could be infected remotely, without the need for a device to be physically added to the network.

Kumar et al. [89] propose a system for early detection of IoT botnets. The authors analysed traffic generated from the *Mirai* malware, particularly during the initial scanning stage, and used it to identify specific signatures which can be used to detect the presence of malware in IoT devices. From this they proposed an algorithm based on a novel two dimensional sampling approach which aims to detect individual bots, rather than the botnet network itself. This approach reduced the computational requirements of the detection mechanism, since only a sample of packets used in the scanning process were analysed. To avoid missing key packets, the detection mechanism was optimised

[1] https://tools.ietf.org/id/draft-ietf-opsawg-mud-22.html#rfc.section.1
[2] https://openwrt.org/

so easily infected devices, categorised as (*vulnerable*), and more difficult to infect devices, categorised as (*non-vulnerable*), were sampled differently. A sampling frequency was specified, which determined the number of devices in each category to be sampled in a defined time period. This allowed more of the vulnerable IoT devices to be sampled. Once the sampling frequency had been fined tuned, the system appeared to perform well, although a number of assumptions were required. Firstly, despite the optimisation, the authors still concluded the algorithm required to be run on a special bot detection device, with sufficient processing power and memory. This could limit the application and scalability of the system in large ISP networks. Secondly, there was an assumption that ISPs would have knowledge of which devices were regarded as vulnerable and non-vulnerable. However, it wasn't clear how this would be achieved, other than from a rudimentary observation that devices in homes would be most vulnerable. The authors therefore claim that IoT devices in enterprise, industrial or governmental networks are less vulnerable, which has not been proven.

Behavioural analysis was found to be another popular method for detecting malware. In [90] Said et al. explored two malware detection techniques used to detect IoT malware. First, syntactic analysis was used with string-based signatures. Rules were created based on syntactic properties (file size and string values) and tested using a malware detection tool (*Yara*[1]). Secondly, behavioural analysis was performed by creating behaviour signatures from samples of the *Mirai* malware. For example, system calls made by the malware when interacting with a host system were recorded, and used as a signature to identify a pattern of behaviour. Both techniques were tested using 500 unique Mirai samples, previously captured using a honeypot. The results demonstrated behavioural analysis to be the more effective technique, however, the authors ultimately recommended a hybrid of the two techniques should be deployed for maximum performance. While the results were promising, the authors did recognise that the techniques had limitations. Indeed, it would be quite easy to evade detection from the syntactic technique by simply changing the size of the binary so that the signature threshold value was not invoked. Secondly, the behavioural analysis technique relies heavily on samples of existing malware, therefore, could prove less effective for new or variants of existing malware.

Anomaly-based Detection. Behavioural analysis can also be used for anomaly-based detection. In [91] Zhao et al. present a method of detecting botnet activity using traffic behaviour analysis and flow intervals. The authors explored the feasibility of detecting bot activity during both the C&C and attack phases based on the observation of its network flow characteristics for specific time intervals. The system was trained

[1] http://virustotal.github. io/yara/

using a subset of the ISCX 2012 IDS dataset, and a second dataset containing the *Zeus* botnet. The system was then tested using two novel (at the time) botnets: *Weasel* and *BlackEnergy*, both capable of performing DDoS attacks. A decision tree classifier was used successfully to detect bot activity with 100% accuracy returned for both malware. However, while the BlackEnergy returned a 0% false positive rate, the Weasel malware returned 82% false positive rate. The significant difference in false positive rates was likely due to the fact the BlackEnergy dataset consisted of only malicious traffic, while the Weasel dataset contained a combination of both normal and malicious traffic. Indeed, the author recognised this limitation suggesting future iterations could be trained using a sample of normal traffic from the target network. While this could certainly improve detection, it would dilute the effectiveness of the system.

Recently, deep learning has also been increasingly applied to the detection of anomalies, where Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN) or unsupervised learning such as auto-encoders have been successfully used. In [92] Meidan et al. extract behaviour features of IoT traffic in a network and use deep autoencoders to detect malicious traffic from the *Mirai* and *Bashlite* IoT malware. In total, twenty three features were collected and used to train a neural network using clean uninfected traffic. Since the autoencoder was only trained on benign traffic, it was able to reconstruct normal observations (normal traffic), however, failed to reconstruct abnormal observations (malicious traffic), classifying these as anomalous. The accuracy of the detection engine was tested by infecting common IoT devices (doorbell, security camera, baby monitor) and measuring the mean TPR, FPR and detection time. Finally, the results were compared to those of three other algorithms: Support Vector Machine (SVM), local outlier factor (LOF) and IsolationForest. Overall, accuracy was very good although one device appeared to be more difficult to detect, suggesting further refinement when capturing normal traffic behaviour is required. Similarly, Kumar et al. [93] propose a system to detect and mitigate attacks from IoT botnets. The system was composed of three components: botnet detection using a sparse autoencoder, cryptomining detection algorithm, and a honeypot used as a decoy IoT device. The neural network-based anomaly detector tested the accuracy of several machine learning models to detect four variants of *Mirai* (*FBOT, ARIS, EXIENDO* and *APEP*). The system also attempted to detect the presence of crypomining by monitoring CPU usage of IoT devices, and identifying anomalies when devices exhibited high CPU and memory utilisation, as a result of the malware running complex crypto computations. Finally, the honeypot was deployed to attract the malware and control the propagation of the botnet. The system returned high detection accuracy, although the accuracy of the support vector machine algorithm was found to be considerably lower. It was also not clear whether the results represented mean values for the four malware variants.

In addition, it was also not clear how the honeypot had prevented the spread of the malware, since the malware would continue scanning and connecting to other devices in the network, beyond the failed login attempt in the honeypot.

In [94] the authors proposed a system based on a convolutional neural network (CNN). Rather than extract features from traffic, the raw traffic data was taken as images, and the CNN used to perform image classification. Traffic was preprocessed, anonymising, and sanitising images before they were converted to IDX format, ready to be ingested into the CNN. Precision, recall and F1 scores were good, however, the authors acknowledged that the system was limited, since it was only able to classify known malware, and was not tested for unknown variants.

Recurrent Neural Networks (RNN) were used in [95, 96] to detect anomalies in an existing dataset. Since DDoS attacks flood a network with similar packets, the appeal of using a RNN is their ability to connect previous information to the present [97], which in the case of a flood attack would mean using previous packets to inform the understanding of the present packet. In [96] a RNN was used, with forty one features transformed to numeric values, before being normalised [0,1]. A total of one hundred twenty two input nodes, and two output nodes were created. The results of the model were measured for accuracy, and compared to other machine learning methods. The results produced were good, maintaining high accuracy, however, training time was higher when compared to other algorithms, suggesting further fine tuning, and feature reduction should be considered. The increased training time could also be due to the method used by RNN to achieve information persistence (long-term dependencies). Although they contain loops allowing information to persist, as the gap between previous information and the present state grows, RNNs become less effective at learning to connect the information. As a result some within the research community have investigated the use of Long short-term memory (LSTM) [97, 98, 99], a special kind of RNN, capable of learning long-term dependencies [98]. Results were promising and demonstrated the potential for detecting, DoS, DDoS and IRC-bot communications in flows of network traffic. In [100] Long et al. propose a deep learning intrusion detection system using word embedding and a Long short-term memory (LSTM) recurrent neural network. The use of word embedding was used to address the challenge of training a model with high dimensional features. Dimensions of features were therefore reduced, while keeping the similarity relationships in semantics and syntax. The Recurrent Neural Layers helped trace the history from previous network packets, and a softmax classifier was used to determine whether the input traffic was normal or malware. The results obtained were good demonstrating that the use of word embedding for network intrusion detection has promise, and could be applied to other areas such as the IoT.

Gu et al. [101] present *BotSniffer* a network-based anomaly detection system built on top of *Snort* to identify botnet C&C channels. In their previous work [87], the authors presented a system which used an extensive set of malware-specific signatures to detect a range of IRC-based botnets. In this research, the system does not require *a priori* knowledge of signatures or C&C server addresses. Instead, spatial-temporal correlation is used since pre-programmed C&C communications and activities will likely display similarities for bots belonging to the same network. Anomaly-based detection algorithms were developed to detect both IRC and HTTP based C&C which were able to detect activity even when a low number of bots were present and the C&C communication was encrypted. The presented results show the system to be effective, detecting all tested botnets and generating very few false positives. Despite this a number of limitations are present, many of which are acknowledged by the author. Firstly, the system uses a whitelist of addresses identified as normal, which is hard coded into the system and could easily be evaded if known. Secondly, the system was tested using standard protocols for C&C such as IRC, however botnets which use bespoke protocols could possibly again evade detection. Finally, the authors did not test the scalability of the system, which could pose an issue in networks with multiple ingress/egress points. In [102] the authors recognised some limitations of their previous work and proposed *BotMiner*. The research acknowledged that bots were evolving and being developed to use different C&C protocols and structures, should as distributed Peer to Peer (P2P). Their aim with *BotMiner* was to develop a system independent of the protocol, structure and infection model, that should also not require *a priori* knowledge of signatures or C&C server addresses. The system is built on the understanding that bots within the same botnet will exhibit similar C&C communications and malicious activity patterns. Thus, the architecture of the system is built on two planes: *A-Plane* for monitoring and clustering activity patterns, and *C-Plane* for monitoring and clustering communications patterns. Clustering results from each plane are sent for cross-plane correlation to find intersections between the two, which may suggest evidence of a host being part of a botnet. The results demonstrated excellent detection accuracy, and was presented as an improvement over their previous work. However, as with any IDS the system does have limitations allowing bots to possibly evade detection. For example, both planes (A&C) likely rely on consistency in activity and C&C communications in order to successfully detect the bot presence. If a bot was developed to exhibit irregular C&C communications and malicious activity patterns, the system may struggle to identify and cluster the activity. This could potentially impact the accuracy of detection.

Nobakht et al. [103] proposed an intrusion detection system (*IoT-IDM*), to provide network level protection for smart devices deployed in home environments. The system utilised a hybrid approach capable of using learned signature patterns of known attacks, and customised machine learning techniques. The system also utilised SDN technology, to provide network visibility using OpenFlow, and security management was provided by a third entity as Security as a Service (SaaS). Here the SaaS provider maintains the database of IoT devices and known attacks on behalf of home networks, and provides updates to repository. To detect possible intruders a predictive model, capable of distinguishing between legitimate access and an attack was used. Logistic regression and Support Vector Machines (SVM) returned high accuracy, however the author acknowledged the limitations of the system. The system relied on a third party to maintain a database of known attack patterns which were used to train the detection modules. It is unlikely that this database could be kept sufficiently up to date, and even if this was possible, the delay between new attacks patterns being learnt and homes being updated, would leave them open to zero day attacks.

Specification-based Detection. Adat et al. [104] propose a DDoS mitigation framework to protect resource constrained IoT devices. Specifically, they produced an algorithm to defend against denial of service attacks based on resource exhaustion. The proposed algorithm consisted of two modules: an analysis module to classify incoming traffic as suspicious or normal, and a monitoring module to categorise the attacks as denial of service (DoS) or distributed denial of service (DDoS). Detection was performed at the border router and packets initially checked against a blacklist, with non-black list IP addresses allowed through. The system also used *bit rate* as a metric to check for normal or anomalous packets. The presented results were good, however it was not clear which DDoS attacks were performed. Also, the system was only tested in a simulated environment, therefore results are limited and accuracy unknown in a live setting.

Threshold values are often specified and used as triggers to identify potential malicious activity. In [105] an IDS was developed using Suricata[1], a free and open source real time threat engine capable of inline intrusion prevention (IPS), network security monitoring (NSM) and offline pcap processing. They developed a novel security architecture for detecting DoS attacks in 6LoWPAN Wireless Sensor Networks (WSN). Suricata's threshold rules were used to trigger an alert when matched packets in a UDP flooding attack, exceeded a defined threshold. The system proved capable of detecting DoS attacks, but was limited since rules could be easily bypassed, and the rules would require to be constantly updated. The authors extended their work in [106] where alerts

[1] https://suricata-ids.org/

were now sent to a Security Information & Event Management (SIEM), however, the limitations of their previous work appear to have been carried forward, and therefore remain. Ahmed et al. [107] also used threshold values to identify malicious activity by proposing a novel blockchain based solution for detecting IoT botnets. IoT devices in the network connect to the Internet via Autonomous System (AS), which are responsible for forwarding device traffic, and maintaining four lists of IP addresses: *Blacklist, Whitelist, Suspected Attacker List* and *Possible Victim List*. Multiple AS exist and are connected together via the blockchain, to allow each AS to share their IP lists with other AS in the network. In addition, every AS monitors the behaviour of the devices connected to it, and compares their behaviour to pre-defined threshold values. If a threshold value is exceeded the device is added to the blacklist, and then shared throughout the network. Known trusted devices are added to the whitelist, and new IoT devices are added to the suspected attacker list. Once identified as trusted, the new IoT devices are moved to the whitelist. The system was evaluated in a simulated environment, and once fine tuned, returned true detection rates of 95%. While the system proposed an interesting and novel approach to botnet detection, it was only tested in a simulated environment. Similar approaches have also been proposed in [108, 109], however the feasibility of maintaining a Proof of Work blockchain in a real environment, for the purpose proposed in the studies is debatable, since it would likely require a large number of nodes to host a live chain, and could consume a considerable amount of electricity. Bhardwaj et al. [110] also proposed using a threshold value to detect malicious activity utilising edge computing to deploy functions at the edge of a network to gather information about incoming traffic. By sending the gathered information via a fast path to an internal detection service, they report up to ten times faster detection of DDoS attacks in the IoT. To detect an IoT DDoS attack, the edge system records the arrival time of a packet and compares the timestamp with the arrival time of the previous packet. The time difference is referred to as the *inter-packet spacing* and is checked against expected values for HTTP request rate and UDP transmission rate. If the inter-packet spacing drops below a specified threshold, an alert is raised. The use of computational resources at the edge of a network, to monitor traffic and accelerate detection, is an interesting and novel approach, however, has some limitations. For example, in large networks with multiple points of ingress/egress, the edge service would need to be deployed in multiple locations. Additionally, routing traffic internally may have an impact on the speed at which detection alerts can be sent to the internal detection system.

Lee at al. [111] present a lightweight intrusion detection system for detecting DoS attacks in 6LoWPAN WSNs. Interestingly, they used the energy consumption of a node to determine if activity within the network should be classified as normal or malicious.

They used a DoS attack to force a target node to be in a constant state of receive, causing it to increase the energy it uses. Nodes with irregular energy consumption were used to identify malicious activity. The work was similar to the approach used in [112], however, Lee's proposed detection scheme used both route-over and mesh under routing schemes to detect malicious nodes. The detection scheme did not take into account the position of a node within the network, and therefore genuine differences in energy consumption that may exist. For example, a sensor node that is locate near to a sink node (one hop away from it) will likely consume energy differently than other nodes because they have to receive and transmit packets from and to other nodes.

2.2 Situational Awareness

A multitude of definitions and interpretations exist to explain the concept of *Situational Awareness* [113]. At the most basic level, the term is understood to mean the process of acquiring knowledge about the things going on around us [114]. The concept and use of the term is thought to have been birthed within a military context, when it was used to describe the theory of military grouping. Alberts et al. [115] describe situational awareness as:

> "*the awareness of a situation that exists in part or all of the battlespace at a particular point in time. In some instances, information on the trajectory of events that preceded the current situation may be of interest, as well as insight into how the situation is likely to unfold. The components of a situation include missions and constraints on missions (e.g. ROE), capabilities and intentions of relevant forces, and key attributes of the environment.* "

Although clearly a definition contextualised to provide meaning and understanding in military planning, if slightly sanitised the definition can be used to provide meaning to a wider range of contexts.

> "*the awareness of a situation that exists ... [A]t a particular point in time. In some instances, information on the trajectory of events that preceded the current situation may be of interest, as well as insight into how the situation is likely to unfold ....* "[1]

This new sanitised definition provides a clearer understanding of the concept and aids the identification of correlations between the theories and models presented in the next section. It is important to note that in both definitions the emphasis on time remains, and is used to allow anticipation of the future, given the current situation. Tadda et

[1] Note: Ellipsis points used to represent an omission from the original quote

al. [113] highlight the importance of time in the original definition explaining that it allows the use of past experience and knowledge to identify, analyse, and understand the current situation and the projection of possible futures. They contend that this empowers a person to maintain awareness, make decisions, and take action to influence the environment in the future, resulting in an endless cycle of update, decision and action.

2.2.1 Situational Awareness (SA) Theories and Models

In [116] Stanton et al. suggest that three main theories exist to explain situational awareness: Endsley's Three level model [117], Bedny et al. Interactive Sub-systems model [118], and Smith et al. Perceptual Cycle [119]. Stanton suggests that Endsley's model focuses more on the perception and understanding of the environment, with less focus placed on projecting the future. Smith places the focus on the interaction between the person and the world. Finally, Bedny emphasises the role of reflection in situational awareness. When applied to the cyber domain it should be noted that any proposed model should not promote a serial process, but rather a parallel one [113]. Elements in a cyber environment are dynamic and constantly changing, therefore, a model is required which provides continuous updates to and from each layer. Of the three models presented, Endsley's model is most suited to the research in this thesis since it fits this criteria, whereas the two competing models place unequal emphasis in a single area. Endley's model will now be discussed further.

From a review of the existing literature, definitions of situational awareness (SA) can often be traced back to the seminal work presented by Endsley in [117]. A study was presented which investigated if enhancing SA in aircraft pilots could increase their likelihood of making optimal decisions in dynamic situations. In [1] 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 [1] 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 composed of three levels, namely *Perception*, *Comprehension* and *Projection* which combine and contribute to achieving a level of awareness in a given situation. At the first level of the model (*Perception*), Endsley described how a pilot would perceive elements in the environment, such as aircraft instrumentation, control of the aircraft and other aircraft in the vicinity. The model emphasises that at this level no interpretation of information

is required, rather it is collected from a range of sources [116]. At the second level of the model (*Comprehension*), the pilot synthesises all the information collected at the previous stage to understand the significance of each element and its relationship to other elements. For example, the amount of time it will take to travel a certain distance, and if the aircraft has sufficient fuel to make the journey. Finally, at the third level (*Projection*) the pilot uses the synthesised information from the previous level to predict how the elements will affect the future, predict likely scenarios that may occur, and make the appropriate decision at that moment in time.

Again, if the definition provided by Endsley is slightly sanitised it can be used to provide meaning to a wider range of contexts, and 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.

Endsley highlighted the importance of system design when trying to improve situational awareness. She proposed a set of interface design guidelines, which were succinctly summarised by Stanton et al. [116] as:

1. Reduce the requirements for people to make calculations.
2. Present data in a manner that makes level 2 SA (understanding) and level 3 SA (projection) easier.
3. Organise information in a manner that is consistent with the persons goals.
4. Indicators of the current mode or status of the system can help to cue the appropriate situational awareness.
5. Critical cues should be provided to capture attention during critical events.
6. Global situational awareness is supported by providing an overview of the situation across the goals of the operator.
7. System-generated support for projection of future events and states will support level 3 SA.

8. System design should be multi-modal and present data from different sources together rather than sequentially in order to support parallel processing of information

In [120] McGuiness et al. extended Endsley'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 [113]. Tadda et al. [113] suggest that both models represent a parallel process not serial. They contest that information at each level of the model continuously updates the other levels, moving information and understanding throughout the levels to achieve a state of situational awareness in the moment.

2.2.2 Cyber Situational Awareness (CSA)

When applied to the Cyber domain, Cyber Situational Awareness (CSA) 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 [20]. As previously mentioned, network security monitoring can be traced back to Anderson's [75] seminal work analysing security logs for anomalies. Denning [76] continued this work producing a framework which 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 [121]. 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 [113] Tadda et al. 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 [122] Onwubiko identifies the functional attributes of situational awareness for network

and 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 [2] the author extended the work and presented an adapted version of Endsley's SA reference model [1]. The model incorporated Endsley's initial levels *Perception, Comprehension and Projection* and also the fourth level *Resolution* proposed by McGuiness et al. [120]. 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:

0. *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.

1. *Perceive:* use of individual tool-kits 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*.

2. *Comprehend:* use of analysis tools and techniques to continually analyse 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.

3. *Project:* analysed intelligence once comprehended, can be used to predict future events and situations. Performed as a real-time continuous process, allows possible mitigation's against threats to be recommended.

4. *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, however, form a good basis for assessing and enabling the application of SA in the Cyber domain.

2.3 Conversational Agents

A conversational agent is a software agent that interacts with users using natural language [123]. Often referred to as a *Chatbot*, *Bot*, *Virtual Assistant* or *Digital Assistant*, they use a dialogue system to assist users to complete tasks through aural or verbal interactions. Indeed, Furey et al. [124], suggest that voice controlled IoT devices have become ubiquitous in homes and offer individuals many convenient and entertaining features.

The history of conversational agents is often traced back to 1950 when Alan Turing, a pioneer for computer science, posed the question "*Can machines think?*" [125]. In doing so, Turing was attempting to address the problem of artificial intelligence, and define a standard for machine intelligence. Turing acknowledged that this question would be difficult to answer, so refined it to be less ambiguous "*Can a computer communicate in a way indistinguishable from a human?*" [126]. In the same article he outlined a test which could be used to answer this question, and determine in a conversation whether a person was speaking to a human or a computer. The idea was simple: for a machine to pass the *Turing Test*, it must exhibit behaviour indistinguishable from that of a human being. In many ways, this was the beginning of conversational agents and artificial intelligence, mapping a path for developers to create the ultimate conversational experience. Figure 2.4 shows a timeline of prominent conversational agents since Turing published in article in Computing Machinery and Intelligence [125].

Eliza. The first notable agent (*Eliza*)[1] was developed in 1966 by Joseph Weizenbaum [127]. The program was able to convince some users they were talking with a human, however, ultimately failed to pass the Turing test [126]. The program provided a foundation for future agents, adopting the use of keywords, specific phrases, and pre-programmed responses.

Parry. Six years after Eliza, *Parry* was developed. A conversational agent taking on the persona of a person exhibiting traits of schizophrenia. Developed by psychiatrist Kenneth Colby, Parry was programmed to misinterpret a users answers, suspecting the user was concealing hidden motives, and consistently deflected their inquiries to simulate paranoid thinking. Parry was interviewed by several expert mental health professionals, who could not distinguish its linguistic behaviour from that of paranoid patients [128].

Jabberwacky. Although many computer scientists were intrigued by Turing's question, it took sixteen years for the next agent of note to appear. *Jabberwacky*[2] was

[1] https://www.masswerk.at/elizabot/
[2] http://www.jabberwacky.com/j2about

developed in late 1980s, and was designed to simulate natural human conversations in an entertaining manner. The agent worked by storing important sections of conversations and used contextual pattern matching techniques to find the most appropriate response.

Dr Sabaitso. A milestone in the development of conversational agents occurred in 1992, with the introduction of the first agent able to synthesise speech. Powered by a Creative Labs *Sound Blaster* sound card, it communicated aurally with users, giving the appearance of more human features than its predecessors [126].

Alice. Three years later, Richard Wallace developed the Artificial Linguistic Internet Computer Entity, more commonly known as *Alice*. Created in AIML (Artificial Intelligence Markup Language) the architecture split the chatbot engine, and language knowledge model, to provide extensibility for alternative language knowledge models to be plugged in [123]. Although unable to pass the Turing test, the agent was highly regarded, receiving a number of awards for being the most advanced agent of its time.

SmarterChild & Watson. The early 2000s saw a number of agents introduced, with different emphasis. Notably, *SmarterChild* an intelligent agent heavily used by users of *MSN Messenger* and *AOL*. The agent was designed to be a fun, personalised experience, and is regarded by many to be the precursor to Apple's *Siri* conversational agent. Around the same time IBM's *Watson* also appeared, built to compete at a human champion level on the American TV quiz show *Jeopardy*, beating two former champions [129]. In addition to its question-answering accuracy, the agent was also designed with speed, confidence estimation, and clue selection built in to improve its chances of winning.

Development of conversational agents continued to grow steadily, due in part to Turing's original questions, and also inspired by a global competition formed in 1990, run by *ASIB - The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*[1]. The *Loebner Prize* was sponsored by Hugh Loebner, an American inventor, with the intention of implementing the Turing test and finding the first computer that could generate responses indistinguishable from those of a human [130]. A second notable prize was added in 2017 by Amazon who have fully embraced the voice computing era stating "*The way humans interact with machines is at an inflection point and conversational artificial intelligence (AI) is at the center of the transformation*". Sharing similarities with the Loebner prize, the *Alexa Prize* was established to accelerate the field of conversational AI, providing a platform for university students to showcase their skills. The main difference between the two major competitions is that in the

[1] https://aisb.org.uk/

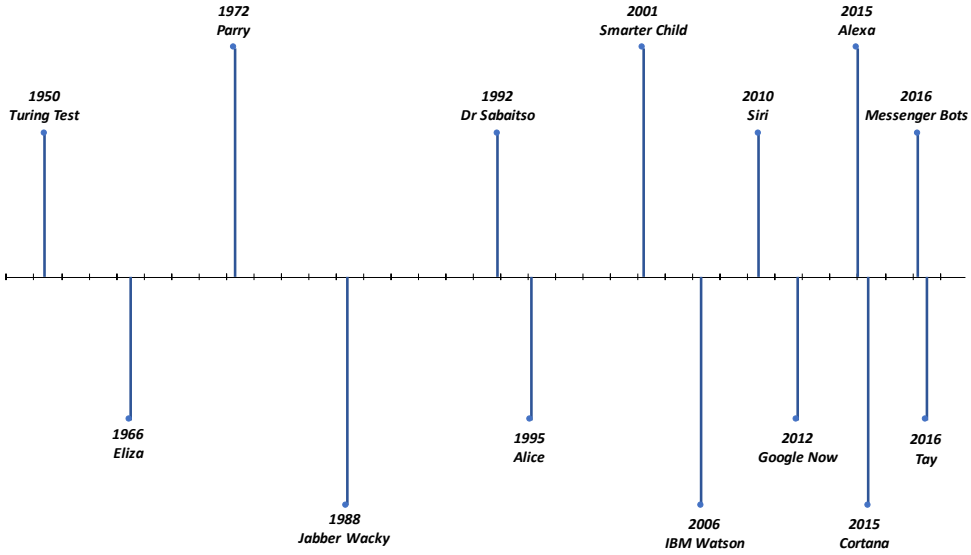


Figure 2.4: Conversational Agent Timeline

Alexa prize the developed agents do not try to pass as human, but rather assume the role of an assistant, focusing on conversing through fluent and enjoyable interactions [131].

The agents discussed so far provided the foundation for arguably the most important decade in the history of conversational agents. The decade between 2010-2020 has produced seismic advances in conversational agent development, fuelled in large part by artificial intelligence. James Vlahos suggests that while IBM may have dominated the mainframe era; Microsoft the desktop computer; Google monopolised Internet search; and Apple and Facebook revolutionised mobile computing, we are now entering a new era, the era of voice computing [131]. One in which personified AI assumes the role of helper, oracle and friend, to enable us to control any, and every piece of technology. The growing popularity of conversational agents cannot be understated, with agents 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. Simply put, voice controlled agents are considered the realisation of science fiction, the dream of interacting with computers by talking to them [132]

Siri. This new era of voice computing was started at an event in 2011 called *Let's Talk iPhone*, when Apple released the agent it had been working on called *Siri*. Sadly, Apple CEO Steve Jobs didn't get to see the evolution of voice computing, dying from

pancreatic cancer the day after Siri was released. The inventor of Siri, Adam Cheyer had worked on fifty predecessors of Siri, before finally releasing a personable, conversational being that interacted in natural language, and could connect with other programs and services to retrieve information or accomplish tasks [131]. Built on a modular and expandable architecture, the agent was taught to grasp the overall intent of a given utterance (speech request) rather than learn all the rules of language such as nouns and verbs, parsing every single word which is a labour intensive process.

Google Now (later known as Assistant). Google quickly followed in 2012 with *Google Now*, a voice activated search facility, which was platform agnostic and allowed users to search the web and perform a variety of tasks such as scheduling events and posting to social media. Branding was later phased out, and functionality incorporated into the newly branded *Google Assistant* in 2016 with the launch of *Google Home*. A smart speaker capable of engaging in conversation with users and assisting the automation of tasks.

Alexa. In 2011, Amazon CEO Jeff Bezos, an ardent Star Trek fan, charged his chief technical advisor Greg Hart with the task of building the first voice only computer. *Project Doppler*, a secret Amazon project exploring conversational AI was established, and acquired three companies *Yap*, *Evi* and Ivona which specialised in speech recognition [131]. The acquired technology was used to create a small cylindrical device with six directional microphones, capable of being triggered using a *Wake Word*, initially known as *Flash*, but later launched in 2015 as *Echo*. Powered by a conversational agent known as *Alexa*, the device was able to fully converse with a user and work on their behalf to retrieve information from the Internet and automate a variety of tasks such as playing music and ordering a pizza.

Cortana. A noticeable absentee from the list so far is Microsoft. In 2015, they launched *Cortana* their first foray into the field of voice computing and conversational agents. Development initially start in 2009, under Zig Serafin and Larry Heck, and was designed as an assistant to help schedule a daily routine and find information on the Internet. Although perhaps late to the game, Microsoft have been quick to accelerate their exploration of conversational agents with *XiaoIce*. Following the success of Mattel's *Hello Barbie*, Microsoft have been quick to explore the use of agents for affective computing, exploring the development of emotional connections between a user and an agent [131]. XiaoIce was designed as a "*Friend*" to teenagers, powered with emotional intelligence, recognising not only the expressed request, but detecting the sentiment, in order to respond with the correct emotional response. Indeed, personification - the attribution of a personal nature or human characteristics to something inanimate, is a growing field of research within voice computing and conversational agents, generating

interesting research such as [133].

Tay. While great advancements in agent design have been experienced recently, developers are still cautious about how much autonomy should be granted to an agent. A case in point is *Tay*, a Twitter based agent developed by Microsoft and powered by unsupervised machine learning. Launched in 2016, the agent caused controversy when it began to post tweets using inflammatory and offensive language [131, 134]. The agent had been targeted by mischievous users, who were quick to test its ability to learn by engaging it in controversial conversations, which it duly learnt and included in subsequent responses.

Further concerns exist relating to security and privacy, in particular in voice activated agents such as Google Assistant and Amazon's Alexa [131, 124]. Although devices such as Amazon's Echo have been shown to support a number of security features, preservation of privacy is reliant on the user implementing the measures in order to protect their own data. In addition, anecdotal evidence by Artem Russakovskii, highlights the issue of "accidental recordings", something Google acknowledged as phantom events, and assured had been fixed. The question of privacy still largely remains, and continues to be a topic of interest for future researchers.

Despite these concerns, research into the use of conversational agents is growing, and producing some very promising applications and use-cases. Indeed, Io et al. [135] believe more businesses should explore the use of agents, and suggest the benefits could expand beyond business to education, psychology and linguistics. Paikari et al. [136] also see the potential of agents suggesting that even though at present the number of agents using voice input and spoken output is still relatively small, this method of interaction will become increasingly more prevalent. However, not all researchers agree with these sentiments, suggesting voice controlled agents may be limited to simple tasks, and struggle to execute more complex tasks, such as searching databases and requesting inter library loans [132].

Rajalakshmi et al. [137] successfully demonstrated the potential of conversational agents by presenting a system using Amazon's *Alexa* and *Node-Red*, a simple and powerful automation platform, to interconnect and control numerous IoT devices. The agent provided the ability to switch smart lightbulbs on/off, monitor iPhone statistics, and used voice commands to control a heater. The research offered a lot of promise in the area, however a lack of detail made it difficult to fully assess the functionality.

Yue et al. [138] proposed a smart home system using the *Reverb* and *Telegram* mobile apps to control smart appliances in the home. The reverb app was used to send voice commands to the Alexa Voice Service in AWS, which interfaced with a local raspberry

pi, to switch a device on/off. The Telegram app was used to send commands via text, and perform similar tasks. Functionality was limited, but returned positive results, demonstrating good promise in this area.

Solorio et al. [139] proposed a voice activated semi-autonomous utility vehicle. The authors successfully managed to adapt a lawnmower, which could be controlled using the Alexa Voice Service in AWS, moving in four directions, at five different speeds. The adapted vehicle was fitted with ultrasonic sensors for obstacle avoidance, and testing returned impressive results. The research clearly demonstrated the potential for future applications in the area of automated mowing and transportation.

In [140] the authors proposed a system to automate detection notifications from a security camera. The *Sbot* agent was deployed on Facebook Messenger and utilised a backend Human Detection Server, linked to a smart camera, to monitor zones within a property. When a human passed in front of the camera, the detection engine generated a notification message, and used the agent to deliver the information to the user via Facebook. Results showed the system to be 95.7% accurate in an indoor environment, and 91.2% accurate outdoors. The dialogue used by the agent was quite limited, but did demonstrate the potential of using agents to assist users secure a property.

The next generation of conversational agents were proposed in [141]. A multi-modal dialogue system was developed to combine multiple user input modes, such as speech, touch and verbal/non verbal gestures. The authors proposed a system which used a camera and Microsoft Kinect device to receive speech and gesture input commands, which were processed and stored in a knowledgebase. The application of the system was unclear, however future use cases were suggested including, educational assistance, robotics and home automation.

Reis et al. [142] proposed the use of conversational agents to assist the elderly. The authors suggested that agents 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 to improve the issue of social isolation. 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, conversational agents could be used to address.

Kerly et al. [143] investigated the use of conversational agents in E-Learning, presenting two examples of agents used to provide learning support and self assessment. They argued that the use of natural language agents in education allowed a user to focus their cognitive efforts on learning the task at hand, rather than expending energy on

the communication medium being used. *CALMsystem* used an agent to assist a user to assess their understanding of a topic using a natural conversation. The second agent *TeachBot* was designed to assist a user to consolidate their knowledge of a subject and provide advice about completing tasks, such as writing an essay. In post trial results users reported finding the systems helpful, easy to use and fun.

Shepherd et al.[144, 145] explored the use of avatars, to investigate if affective feedback could be used to improve awareness of end-user security. Specifically, if human-like agents could use subtle facial cues to improve awareness of risks, and encourage users to in engage in more secure behaviour. Previous studies [146] had identified six basic emotions that could be used to provide affective feedback in a given situation: happiness, anger, sadness, fear, disgust, and surprise. Shepherd et al. used happiness and sadness to denote positive and negative feedback accordingly, demonstrating the user human-like avatars to be an effective method of enhancing a users general awareness of security risks. The results presented by Shepherd et al. [145] provide a strong case for exploring if agents with human-like features could be used to address the lack of awareness of threats suggested by Legg [3].

Finally, some interesting research was conducted in [133] which investigated the personification of conversational agents, 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 conversational agents offer a wide range of use cases and applications. The willingness of users to adopt this new method of interacting with devices and information is likely to promote wider use in the future, presenting interesting opportunities to researchers.

2.4 Conclusions

In this chapter, existing literature relevant to the research presented in this thesis, was reviewed. It is clear that the IoT is a novel paradigm with the potential to revolutionise large sections of everyday life. However, the literature also demonstrated that many smart devices found within consumer homes are vulnerable and can be easily leveraged to perform large scale attacks. In particular, the evidence suggested IoT botnets are becoming increasingly more common and sophisticated in their effectiveness and ability to exploit basic security vulnerabilities in consumer IoT devices, and obfuscate their activity.

Intrusion detection was shown to be a topic of growing research, particularly when used to detect malware targeting the IoT. More specifically, when used as a form of passive network monitoring, in which traffic is examined at a packet level and results of the analysis are logged. However, output from *Intrusion Detection Systems* (IDS) is typically designed for expert interpretation, making it difficult for non-technical users to harness the power of these systems. Despite this limitation, an IDS can play an important role in supporting consumers to manage their home networks and smart devices. Since many of the attacks discussed in this chapter often target consumer-level smart products, which commonly lack a screen or user interface, it can be difficult for users to identify signs of infection and be aware of devices that have been compromised. An IDS can, therefore, be used as a background monitoring tool, passing output to a user-friendly front-end system, for interpretation by non-technical users.

The existing research demonstrated *Intrusion Detection Systems* (IDS) to be an effective countermeasure against botnet activity when observing and identifying active attacks and vulnerabilities in network traffic [74]. Of particular interest was research in [100] which presented a LSTM-RNN used in conjunction with word embedding to detect the presence of malware. As discussed, the appeal of using a RNN for malware detection is their ability to connect information from previous packets to the present state and inform the understanding of the present packet. In addition, the use of LSTM-RNN can overcome the problem of long-term dependencies, and when used in conjunction with word embedding can be used to provide a method of extracting semantic meaning from information within a packet. However, the existing research performed detection by aggregating network traffic into flows of communication [97, 98, 99], likely collected using a flow collector such as *NetFlow* or *sFlow*. Since the aim of this research is it to investigate threat detection in consumer IoT networks, it is unlikely that SOHO routers would be equipped to collect flow information, therefore, detection would be required at the packet level. A gap in the literature is therefore identified, concerning the use of LSTM-RNN with word embedding to extract semantic meaning from packets, and

perform deep packet inspection. This will be the topic of investigation in Chapter 4.

A second objective of this thesis is to investigate user awareness and perception of threats within the IoT, in order to promote better situational awareness of risks relating to security and privacy. A number of theories and models were discussed [120, 113, 147], however since the aim is to improve awareness, the additional *Resolution* level used in some models is considered out of the scope of this research. Therefore, Endsley's definition of situational awareness is adopted for use in this thesis and is contextualised for Cyber Situational Awareness (CSA). Since Endsley highlighted the importance of system design when trying to improve situational awareness, the interface design guidelines, which were succinctly summarised by Stanton [116] will be considered when this topic is investigated in Chapters 5-7.

The ultimate objective of this thesis is to investigate if conversational agents can be used as a mechanism to improve Cyber Situational Awareness. It is clear from the literature that recent advancements in technology have produced a range of powerful conversational agents, able to fully converse with users to respond to requests or automate a variety of tasks. Numerous interesting examples of agents being used to assist users were identified in the literature. Of particular interest was [141] which presented a multi-modal dialogue system to assist users. Indeed, Amazon also recognise the potential of multi-modal agents, encouraging developers to develop agents[1] that combine other modalities (vision, touch) with voice controlled Alexa agents. Interestingly, Endsley also suggested a multi-modal approach for system design could improve situational awareness [116]. Therefore, having reviewed the existing literature, a gap is identified regarding the use of conversational agents for threat detection and network monitoring, specifically the use of multi-modal agents to aid situational awareness of threats in consumer IoT networks.

In the next chapter, the research methods used throughout this thesis are introduced. Many of the quantitative and qualitative techniques are repeated in several chapters, however, will be presented in the next chapter for ease of reference. Also, by outlining a standard practice and utilising it throughout the thesis, consistency will be established in the research.

[1] https://tinyurl.com/multi-modal

Chapter 3

Methodology

In Chapter 2 relevant literature related to this research was reviewed. The literature was chosen due to its relevance to the central research question addressed in this thesis: namely if situational awareness of threats in the Internet of Things could be improved using conversational agents. This question was further broken down into four sub-questions, each addressing a different but related aspect of the research. This chapter introduces the research methods used throughout this thesis to answer the central research question. Many of the techniques are repeated in several chapters to ensure a level of consistency between studies when addressing the sub-questions in the subsequent chapters.

3.1 Introduction

In this chapter, the techniques used throughout this research are introduced. The central focus of the research question investigates an increasingly important aspect of human behaviour, namely awareness of security threats. In behavioural research empirical investigations can be broadly categorised into three groups: descriptive investigations, relational investigations, and experimental investigations [148]. Descriptive investigations allow a researcher to observe phenomenon, relational investigations enable the identification of relations between multiple factors, and experimental investigations can help determine the casual effect between two factors [149]. The research presented in this thesis combines multiple investigations presented in Chapters 4-7, each of which detail a separate, but related study, addressing one of the four sub research questions outlined in Section 1.2.1. This chapter begins by discussing the adopted research philosophy before continuing to describe methods and data analysis techniques used throughout this thesis.

3.2 Research Philosophy

Each researcher holds a set of beliefs about the world and nature of reality. An understanding of these philosophical perspectives can help a researcher gain an understanding of the wider philosophical perspective underlying the research, and also aid in choosing the appropriate research methods to use [150]. There are four main trends of research philosophy: positivism, interpretivism, pragmatism and realism. Positivist studies incorporate formal propositions, quantifiable measures of variables, hypotheses testing and the drawing of inferences about a phenomenon from the sample to a stated population. In contrast, interpretive studies assume that people create and associate their own subjective meanings as they interact with the world around them, therefore researchers attempt to understand phenomenon through accessing meanings that participants assign to them. [151]. In pragmatism researchers have the freedom to choose the methods, techniques, and procedures that best meet their needs and scientific research aims. Realism is based on the principles of positivist and Interpretivist research philosophies [150]. Chapters 4 - 7 use quantitative approaches since the studies attempt to quantify variables, test hypotheses and use control groups to mitigate confounding variables and enhance scientific rigour. However, in Chapters 5 - 7 qualitative approaches are also adopted to elicit subjective meanings when assessing participants perception and awareness of threats facing smart devices. Therefore, the research contained within this thesis draws upon both a positivist and interpretivist philosophy, since it adopts a mixed methods approach (mix of quantitative and qualitative approaches) [152].

3.3 Research Methods

3.3.1 Sampling Techniques

Specifying the optimal number of participants for a usability study continues to be a topic of hot debate [153]. A widely accepted view within HCI is that five users will find approximately 80% of usability problems and is therefore sufficient for most usability testing [154]. Other research studies however have found that the appropriate number of users is dependent on the size of the project, with seven users being suggested as the optimal number in a small project [155]. For comparative studies where statically significant findings are being sought, a group size of 8-25 participants is typically valid [153]. Participants were required for three of the studies in this thesis and were, therefore, recruited in accordance with these suggestions. In Chapter 5 the aim was to analyse a range of views from a wide audience, to assess how users value and perceive security and privacy in smart devices found within the IoT. An online study was chosen

for format flexibility, and to allow access to different populations [156]. Participants were informed that consent to participate was implied when they decided to engage in the research and complete the online study. In Chapter 6 the aim was to assess the viability of conversational agents for improving awareness of threats facing smart devices. Convenience sampling was employed, with participants selected due to their convenient accessibility, and proximity to the author. Participants were recruited at University events open to the public to ensure a wide range of views were collected and oversampling of a specific demographic did not occur. This seemed appropriate since research into human-computer interaction often uses a more focused population [149]. Since randomisation is an important element of a well-designed experiment [157], participants were therefore randomly assigned to two groups (Control and Intervention). In Chapter 7 the aim was to assess the utility of the conversational agents tested in the previous chapter. In compliance with [155] [153] a suitably sized sample of users was recruited, and again randomly assigned to two groups (Control and Intervention). In addition to protect against confounding the authors in [157] suggest randomising other factors of the study. Here, confounding is defined as when the effect of one factor or treatment cannot be distinguished from that of another factor or treatment [157]. This approach was used in [158] where the researchers found it necessary to randomise the order of the scenarios being used in an experiment. If the order of the scenarios was not randomised, there was a possibility that one scenario could influence the results of the next. To avoid such an occurrence in the studies of this thesis, the order of conversational agent use in study two and three was randomised. Four participants used the aural agent to complete study two, while the remaining four participants used the verbal agent. Agent use for study three was reversed.

3.3.2 Use-Case Development

To evaluate the use of conversational agents for improving situational awareness a set of use-cases were devised, each representing a realistic description of how a user might want to use the conversational agents for monitoring smart device and network activity. These were formulated from existing literature [3], which provided discussion and insight about user engagement with tools designed to improve cyber situational awareness. From the use-cases it was then possible to build a range of different scenarios, which could be used to evaluate the agents in Chapter 6 and 7. The use of scenarios is widely used in a range of areas, such as the military, theatre and software development [159]. The ability to have different scenarios to represent the same use-case was important in order to avoid confounding variables or adding bias into the research. If a participant had seen a scenario in a previous study it may effect their decision making

or performance. Using different scenarios to represent the same use-case ensured confidence in the reliability of the collected data. The developed use-cases used for Chapter 6 are presented in Table 3.1 and the additional use-cases for Chapter 7 in Table 3.2.

Table 3.1: Uses-Cases Chapter 6

	Use Case
uc1	In the first 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.
uc2	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).
uc3	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.
uc4	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 more information about the specific device which may be causing the problem.
uc5	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.
uc6 $\pm$	In the sixth scenario, a user suspects a threat or unusual activity has occurred within the network. They are not sure which smart devices have been active and are possibly compromised. They now query the database to get a list of all active source devices on a specified date.
uc7 $\pm$	In the seventh scenario, previously compromised smart devices have been fixed, but the user would now like to check if network activity has returned back to normal for the last three days. They query the database to check if total network activity has been normal and consistent over the last three days.

$\pm$ *not used in pilot study*

Table 3.2: Additional Uses-Cases Chapter 7

	Use Case
uc8	In the eighth scenario, previously compromised smart devices have been fixed, but the user would now like to monitor the activity level of specific smart device's on the network. They query the database to check if specific smart device activity has been normal and consistent over the last three days.
uc9	In the ninth scenario, previously compromised smart devices have been fixed. They invoked the intent associated with the eigth scenario and found the smart device activity level to be normal. They would now like to be able to identify future risks by monitoring all Smart Device activity levels over several days, to establish normal daily activity levels for each device.

Table 3.3 demonstrates the scenarios used for each study in Chapter 6 and 7, and how they map to the nine use-cases which were developed. The scenarios devised for each use-case in Chapter 6 are described in Appendix D.1, while the scenarios devised for each use-case in Chapter 7 are described in Appendix D.2 - D.5.

Table 3.3: Use-case to Scenario mapping

		uc1	**uc2**	**uc3**	**uc4**	**uc5**	**uc6**	**uc7**	**uc8**	**uc9**	**Append**
Ch 6	**Pilot**	*sc1*	*sc2*	*sc3*	*sc4*	*sc5*	-	-	-	-	D.1
	Main	*sc1*	*sc2*	*sc3*	*sc4*	*sc5*	*sc6*	*sc7*	-	-	D.1
Ch 7	**Study 1**	*sc8*	*sc11*	*sc12*	*sc10*	*sc13*	*sc9*	*sc15*	*sc14*	*sc16*	D.2
	Study 2	*sc17*	*sc20*	*sc21*	*sc19*	*sc22*	*sc18*	*sc24*	*sc23*	*sc25*	D.3
	Study 3	*sc26*	*sc29*	*sc30*	*sc28*	*sc31*	*sc27*	*sc33*	*sc32*	*sc34*	D.4
	Study 4	*sc35*	*sc38*	*sc39*	*sc37*	*sc40*	*sc36*	*sc42*	*sc41*	*sc43*	D.5

Use-Case descriptions found in Tables 3.1 and 3.2

3.3.3 Survey Design and Question selection

Surveys are considered to be one of the most commonly used research methods across all fields of research [149]. They can be an effective tool when a researcher wishes to describe a population or explore and explain behaviours [160]. Survey instruments were used in a number of studies in this thesis (Chapters 5-7) and were delivered

using **LimeSurvey** an online survey platform service[1]. To ensure consistency between studies the general design, structure and rationale for question selection remained the same throughout. In this section these general considerations are discussed and the rationale that informed their adoption highlighted. Deviations from this general design for individual studies are discussed in the study design section for each chapter. For the general survey design, consideration was first given to the overall structure. Trusted guidelines were followed and each survey started with a clear set of instructions [160], detailing how a respondent should interact with the survey. Related questions were grouped together into sections to lower cognitive load on respondents, allowing deeper thought, rather than mentally switching contexts between questions [161]. Debate exists whether demographic questions should be left until the end of the survey, as these are the least interesting [160]. However, as some study hypothesis were dependent on this information, this was considered vital data to collect and was therefore requested at the start of each survey. Finally, it was important to avoid adding bias into responses, so care was taken not to prime questions using biased wording [160]. For example, in Chapters 5-7 the surveys were used to measure participants awareness of threats. To avoid priming, the word "Awareness" was not used and any occurrence of the word was substituted for "Appreciation". By not alerting participants that the study was measuring their awareness of potential threats, participants would be less likely to report inflated confidence levels, which is often found as a result of a phenomenon known as the Hawthorne Effect [162].

3.3.4 Measuring Usability

The performance of anomaly detection techniques are often evaluated from two perspectives; efficiency and effectiveness [163]. The International Organisation for Standardisation (ISO) state "the objective of designing and evaluating systems, products and services for usability is to enable users to achieve goals effectively, efficiently and with satisfaction, taking account of the context of use" [164]. ISO 9241-11 specifically suggests that measures of usability should cover:

- **Effectiveness**: the consciousness of relevant elements in the environment, specifically the status, attributes, and dynamics of elements in relation to the environment.;
- **Efficiency**: the level of resource consumed in performing tasks;
- **Satisfaction**: users subjective reactions to using the system.

[1] https://www.limesurvey.org/

These suggested measures were used to evaluate the viability and utility of the conversational agents in Chapters 6 and 7.

Effectiveness (Accuracy): Precision, Recall and *F*-Measure

Effectiveness is the measure of a systems ability to distinguish between normal and intrusive activities. The most common classification application is binary classification. Within this two-class nature of detection, there are four possible outcomes [165]:

1. **True positive (*tp*):** anomalies that are successfully detected
2. **False positive (*fp*):** normal activities that are incorrectly classified as intrusive
3. **True Negative (*tn*):** normal activities that are successfully classified as normal
4. **False Negative (*fn*):** anomalies that are missed and classified as normal

In Chapter 7, measures of accuracy were used to determine if participants had correctly detected compromised devices using three popular metrics; precision, recall and F Measure. These metrics ignore the normal data that has been correctly classified (tn) and are calculated using:

Precision (P): Defined as the % ratio of the number of true positive (tp) records divided by the sum of true positive (tp) and false positive (fp) classified records.

$$P = \frac{tp}{tp+fp},\ precision \in [0,1]$$

Recall (R): Defined as the % ratio of number of true positive records divided by the sum of true positive and false negative (fn) classified records.

$$R = \frac{tp}{tp+fn},\ recall \in [0,1]$$

F Measure ($\mathbf{F_1}$): Defined as the harmonic mean of precision and recall and represents a balance between them. It is often used to measure the performance of a system when a single number is preferred [165].

$$F_1 = 2 \cdot \frac{P \cdot R}{P+R},\ F\ Measure \in [0,1]$$

Efficiency (Time): Duration of Detection

The second metric used to measure the usability of the conversational agents in Chapter 7 was efficiency [163]. This again complied with the recommendations stated in ISO 9241-11 [164] for specifications of products to ensure quality, safety and efficiency. The efficiency of each of the agents was calculated by measuring how long (in seconds) it took for participants to collect the necessary data about smart device activity and

process this into meaningful information which they could use to determine if a threat had occurred. The Mean, Median and Standard Deviation was calculated for each agent and compared against a baseline **Visual** (Vi) method, to see if the use of the agents had made the process of detecting threats more efficient (quicker). The use of the baseline **Visual** (Vi) method was required to ensure a fair comparison was achieved, since as shown in Chapter 5 without any additional information users find it very difficult to identify an infected smart device and correctly detect threats. If the efficiency (time) of detecting threats using the agents was compared with a scenario that did not provide the user with any additional information, there would be obvious improvement and result in an unfair comparison. Since in Chapter 5 a visual presentation of information was found to improve awareness of threats, this modality was used as a baseline metric for comparison.

Satisfaction: System Usability Score (SUS) Instrument

Originally introduced by John Brooke in 1986, the System Usability Scale (SUS) is a tool used to measure the usability of systems [164]. The tool is technology agnostic, non-proprietary, quick and easy to use, and provides a single score which is easy to understand. Although some excellent alternatives have been created since its release the SUS continues to be a valuable and robust tool and a good choice for general usability practitioners [166]. Usability questionnaires were used in Chapter 7 to assess the utility of each conversational agent (see Appendix G). As suggested by Brooke participants completed a SUS scale for each agent after having had the opportunity to use the system, but before any discussion or debriefing took place [164]. Respondents provided immediate responses to the ten questions and marked the centre point of the scale if they were unable to respond to a particular item. Participants ranked each question on a scale from 1 to 5, based on their level of agreement.

The final SUS score was calculated by first determining the sum of each item. The score contribution for odd numbered questions was adjusted to be the scale position minus 1. The scale contribution for even numbered questions was adjusted to be 5 minus the scale position. Finally, the sum of the scores was multiplied by 2.5 to obtain the overall value of system usability.

3.3.5 Measuring Situational Awareness

The SA model presented by Endsley in [1] 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 composed 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 be made in dynamic situations.

Although extensions to Endsley's model have been proposed [120], since the aim of this thesis is to improve awareness of threats, the additional *resolution* layer was not considered relevant for inclusion. Therefore, to measure participants awareness and confidence to detect threats in Chapters 6 and 7, the SA model presented by Endsley was adopted. The model was used to explore how well participants could assimilate information about events in their environment (*Perception*), synthesise this into a meaningful understanding of the situation (*Comprehension*) and use the knowledge to identify threats in a network (*Projection*). Nine confidence statements were created and mapped to the three levels of Endsley's model as shown in Table 3.4. For each statement, participants were asked to indicate their level of agreement using a five-point Likert scale from *Strongly Disagree (1)* to *Strongly Agree (5)*. In both studies (Chapters 6-7), a Pre-Study Post-Study methodology was used, therefore the participants were asked to provide responses before and after using the conversational agents and differences in their responses compared.

Table 3.4: Cyber Situational Awareness Statements

Five-point Likert Scale from Strongly Disagree (1) to Strongly Agree (5)

		SA Statement
Perception	***pe1***	I am confident I can tell which smart devices are using my home network.
	pe2 ±	I am confident I can tell how often a smart device is communicating on my homework, and how much of the available network bandwidth it is using.
	pe3 ±	I am confident I can tell which smart devices have the highest usage on my home network.
Comprehension	**co1**	I am confident I can tell if my network is experiencing a normal level of device communications and bandwidth usage.
	co2	I am confident I can tell if a smart device is functioning normally.
	co3 ±	I am confident I can tell if a smart device is using my home network more or less than normal.
Projection	**pr1**	I am confident I can tell if an attack has taken place on my home network.
	pr2	I am confident I can tell if a smart device on my home network has been compromised.
	pr3 ±	I am confident I could tell in the future if my home network or smart device had been compromised.

± *not used in Chapter 6 pilot study*

Since data collected to measure situational awareness is ordinal (*Projection*, *Comprehension* and *Projection*) it is unlikely the data will be normally distributed. Therefore, ranked based tests will be used, and described in the next section.

3.4 Data Analysis Techniques

3.4.1 Quantitative Methods

Cronbach's Alpha

Cronbach's alpha (α) is commonly used to measure internal consistency (reliability). The calculated α score would fall in the range [0.0,1.0], and quantifies the degree to which items on an instrument are correlated with one another [167]. It is most commonly used to assess the internal consistency of a questionnaire (or survey) that is made up of multiple Likert-type scales and items. A reliability coefficient of .70 or higher is considered "acceptable" [168]. Cronbach's alpha was used to test the internal consistency of participant SUS responses in Chapter 7, and Situational Awareness scores in Chapters 6 and 7. For Cronbach Alpha tests the α score is reported.

Wilcoxon Signed-Rank

The Wilcoxon Signed-Rank test is the most commonly used non-parametric test for paired data. The test determines whether there is a statistically significant difference in the median of a dependent variable between two related groups. It is the non-parametric equivalent to the paired t-test. As the Wilcoxon Signed-Rank test does not assume normality in the differences of the two related groups, it can be used when this assumption has been violated and the use of a paired t-test is inappropriate[169]. In Chapters 6 and 7 related groups were compared, where participants completed the same survey Pre and Post-Study, therefore Wilcoxon Signed-Rank tests were used for comparisons. For Wilcoxon Signed-Rank tests the Z score and *p-value* are reported, where a $p < .05$ is considered significant.

Chi-Squared

Data is considered categorical when it can be placed into distinct categories rather than being measured as a point on a scale or ranked in order [170]. If the relationship between two variables is of interest, such as whether they are independent or associated, the number of observations simultaneously falling into the categories of two variables can be counted. The chi-square test for independence, also called Pearson's chi-square test or the chi-square test of association, can also be used to discover if there is a relationship between two categorical variables. The assumptions of the Chi-square which must be met include: degrees of freedom greater than one, randomised samples and independent observations [171]. In Chapter 5 detection accuracy was measured and dichotomous data collected. The dependent variable was nominal and the independent variable was ordinal. To check for relationships between the two variables, Chi-squared tests were

performed to check for independence. For Chi-Squared tests the chi-squared statistic value χ^2, degrees of freedom *df*, and *p-value* are reported, where a $p < .05$ is considered significant.

Friedman One-Way Repeated Measure

The Friedman One-Way Repeated Measure is used to test for differences between groups when the dependent variable being measured is either ordinal or continuous, and where the data has violated the assumptions necessary to run the one-way ANOVA with repeated measures [172]. In studies where the normality or variability of the distributions was a concern, a Friedman test was selected to avoid yielding misleading results from an ANOVA [173]. In Chapter 7 efficiency (measured in seconds) of detection using the conversational agents was measured and compared with a baseline visual method. The dependent variable (efficiency) was continuous data therefore a Friedman test was performed to test for differences in the time taken to detect threats. For Friedman tests the chi-squared statistic value χ^2, degrees of freedom *df*, and *p-value* are reported, where a $p < .05$ is considered significant.

3.4.2 Qualitative Methods

Throughout this research a large amount of data was collected in a variety of different forms. Quantitative data was analysed using the methods in the previous section however, the use of survey instruments also resulted in the collection of qualitative data. In Chapters 5-7 the surveys included open-ended questions which enabled a range of rich responses to be collected. In Chapter 7 short structured Post-Study interviews were conducted, resulting in data which required to be transcribed. In Chapter 6 feedback about the agents use and effectiveness was collected and exported in *csv* format. Before any data could be analysed, it was important to first ensure everything was converted into a common format (*csv*), ready for analysis in a qualitative data analysis package called NVivo [1]. To analyse the data a form of thematic analysis known as template analysis was chosen. This method of qualitative analysis is often used when a researcher has some prior understanding of the concepts to be identified. In this approach hierarchical coding is used and a coding template is created, which summarises the data into themes, and organises them in a meaningful and useful manner [174]. The coding template is usually created from a subset of the data, and is later applied to more sections of the dataset, revised and refined. The main procedural steps described in [174, 175] were followed and therefore created a set of *a priori* themes in advance of the coding process. These initial themes were created from the authors prior knowledge of the subject area, and were designated tentative so could be re-defined or

[1] https://www.qsrinternational.com/nvivo/home

removed, if they did not prove to be relevant, useful or appropriate. First, it was important to become familiar with the raw data to be analysed. Common with most thematic approaches the next step mandated a preliminary coding of the data. This involved working through the data and using the *a priori* themes to highlight anything in the text which was thought to be relevant to the underlying research question. Next, the identified themes were used to define an initial *coding template* which provided a good cross-sectional representation of the concepts and ideas in the dataset as a whole. The initial *coding template* was then applied to further sections of the data in an attempt to find potential relevance in the data. With each iteration the *coding template* was refined, adding new themes or modifying existing ones as required, until a *final* template had been defined. Finally, the template was applied to the whole dataset and results were interpreted accordingly. An example coding template is shown in Table 3.5.

Table 3.5: Chapter 6: Example Coding Template

	Theme	Sub-theme	Example Comment
Most Like	Usability	Convenience	"lots of people already own an Alexa so this would be a good way to get people to monitor their smart devices"
		Hands free	"I liked how it was hands free and didn't require a laptop etc"
		Quick	"much quicker than checking each device individually"
		Easy to Use	"easy to get updates about devices"
		Shared Responsibility	"how everybody in the home could share in monitoring their own devices"
	Accessibility	Visually Impaired	"it would be great for anyone visually impaired"
	Interactive	Enjoyable & Fun	"I enjoyed using this technology"
		Educational	"I liked learning new technology"
		New Experience	"I liked playing with an Alexa for the first time"
		Digital Assistant	"I like the idea of making better use of my Alexa to assist me with other tasks rather than just listening to music"
	Awareness	Encouraged better Security	"This would actually convince me to care more about security"
		Improved Security	"non technical people like me can understand it"

3.4.3 Ethical Practice

To collect the necessary data for the research presented in this thesis, the university *Research Governance and Integrity Policy* [1] was followed to establish and promote good ethical practice in the undertaking of the studies. Computing professionals are bound by the general code of ethics outlined by the *British Computer Society (BCS)*, however,

[1] https://www.rgu.ac.uk/files/187/Governance—Ethics/91/Research-Governance-and-Integrity-Policy.pdf

since three of the studies involved human participants, it was also necessary to follow relevant ethical guidelines of the *British Psychological Society (BPS)*. In doing so, care was taken to ensure the guidelines relating to informed consent, information provision and data protection, were followed. The studies in Chapters 5-7 explored aspects of human behaviour and therefore required the involvement of human participants. It was therefore vital that careful consideration was given to the recruitment of participants, the storage of data, and the reporting of results. Recruiting participants and utilising informed consent online can be tricky [176]. Indeed, studies have highlighted concerns over the use of large scale email campaigns to recruit participants [177], arguing if care is not taken participants can engage in an online study without fully understanding what is involved or providing informed consent. Therefore, participants were first recruited from the local student population, and later via only trusted LinkedIn and Facebook accounts, associated with the researcher and School of Computing. In line with the university *Research Governance and Integrity Policy*, at the beginning of each of the three studies, participants indicated their consent to participate by reading a study agreement form (See example in Section 7.2.4) and ticking a box to confirm their understanding and compliance. The form included a description of the study aims, the process of handling data including anonymity of participants and the process for withdrawing from the study. In all three studies no personally identifiable data was explicitly collected, however the use of online survey platforms often (by default) retain information relating to participant IP address and date/time stamps. In line with GDPR guidelines [1] this feature was disabled, therefore preventing the capture of location data and ensuring this information was not collected.

[1] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/

3.5 Conclusions

This chapter introduced the research methods used throughout this thesis. Many of the quantitative and qualitative techniques are repeated in several chapters, however were presented here for ease of reference. By outlining a standard practice and utilising this throughout the thesis, consistency has been established in the research. The next chapter presents the first of the four studies undertaken in this thesis. In this first study, the ability of current detection methods to effectively detect threats found in consumer IoT networks, is explored.

Chapter 4

Botnet Detection in Consumer IoT Networks

Chapter 2 explored how the Internet of Things (IoT) has ushered in an era of increased connectivity [47]. It examined how many of the smart devices found within the IoT are often insecure and vulnerable to misuse. This chapter focuses on one particular threat which has been used to leverage insecure smart devices, in order to perform large scale DDoS attacks on the Internet. The chapter briefly explores the taxonomy of an IoT botnet, using the recently experienced *Mirai* botnet to better understand how infection and spread occur in smart devices and networks. The ability of a current detection method to effectively detect botnet activity is explored, before finally proposing a novel application of deep learning for better detection of botnets found within the IoT.

4.1 Introduction

The focus of this chapter is attempting to answer sub research question SQ1: "*Can current security methods detect the presence of threats within consumer IoT networks ?*". As shown in Chapter 2 the Internet of Things (IoT) has quickly transitioned from a promising future paradigm to a pervasive everyday reality [4]. Billions of smart devices have already been connected to the Internet creating an extensive network of connected 'things'. However, as highlighted, many of these devices are vulnerable, and can used by hackers to perform DDoS attacks against targets on the Internet. This chapter focuses on the detection of botnet activity within consumer IoT networks. A quantitative approach was used to examine how effectively a current detection method can detect the presence of the *mirai* botnet in a sandboxed environment. Subsequently, the use of deep learning and its application to threat detection within the IoT, is explored.

During the undertaking of this study a labelled data set was created and has been made available to the wider research community (See Section 4.5).

4.2 Methodology

4.2.1 Experimental Variables

This study measured an existing security method's ability to detect a known malware threat which targets consumer IoT networks. In addition, a new threat detection method was also developed, and its ability to detect the same threat was assessed. In the process, numerical data was collected and *accuracy* and *loss* metrics were used as dependent variables. The malware used in this study leveraged infected smart devices to perform a range of DDoS attacks. The attacks and C&C messages were used as the independent variables.

4.2.2 Study Design

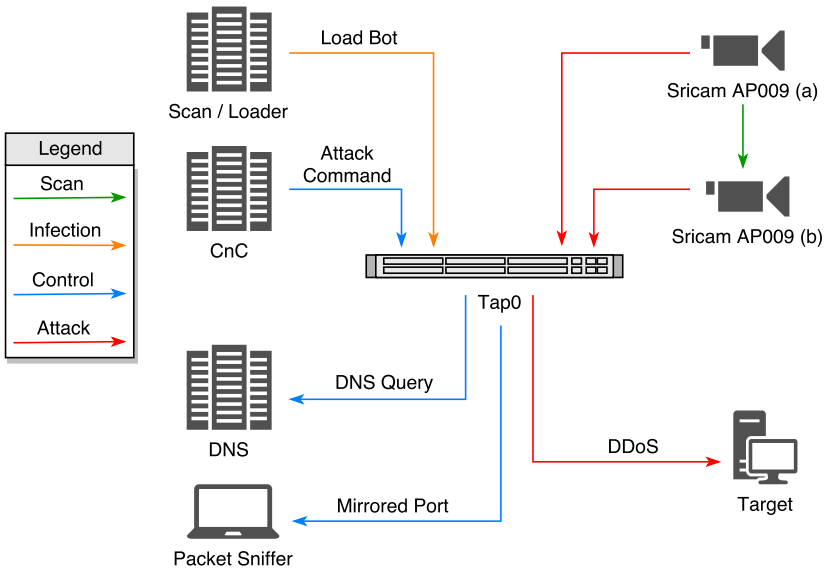


Figure 4.1: Botnet Experimental Setup

A secure sandboxed environment was created as shown in Figure 4.1. This consisted of a command and control C&C server, a Scan/Loader server and an additional utilities server to handle DNS queries and reporting. A soft tap (Tap0) SPAN port was created to mirror all relevant traffic to a packet sniffing device, to capture for later analysis. Two Sricam AP009 IP Cameras running *busybox* utilities were used as bots to attack a target Raspberry Pi. A detailed description of the setup is presented in Appendix A.

The *Mirai* source code was downloaded from GitHub [1]. To ensure a true representation of an infection and attack, amendments to the source code were kept to a minimum, however, some configuration changes were required to comply with ethical and legal regulations. Namely, IP address ranges to be scanned were limited, and DNS queries were directed to a local DNS server in the sand-boxed environment.

4.2.3 Case Study (Mirai): Taxonomy of a Botnet

To foster greater understanding of how botnets target insecure IoT devices, a detailed analysis of the *Mirai* botnet was conducted. The analysis was found to be consistent with other research [10, 178] and is presented below. Figure 4.2 shows the process of infection and propagation method employed by *Mirai*. The *Mirai* infrastructure consists of a command and control (C&C) server, a Scan/Loader server and infected IoT devices known as bots. Infection and propagation occurs by exploiting weak default security credentials found on many IoT devices running *busybox*, an embedded version of Linux. An attacker (botmaster) starts the process by connecting to the Scan/Loader server (*step 1*) and initiating *./loader* to execute the *scanner.c* module, and scan the Internet for vulnerable IoT devices with Telnet services and ports 23 or 2323 open. To avoid detection, 90% of scans use TCP port 23 as their destination, and 10% use port 2323 (*step 2*). Upon detecting a vulnerable device, the malware attempts to brute force a successful login using a list of 62 known default usernames and passwords [10]. If access is successful, the malware runs command */bin/busybox MIRAI*, and waits for reply *MIRAI: applet not found* to confirm the malware is currently not installed on the device. Successful login credentials and device information are sent back to the C&C server, and will be used later by the Scan/Loader server to login and deliver the malware to the vulnerable device (*step 3*). An infect command is sent from the C&C server to the Scan/Loader server containing all necessary information such as login details, IP address, hardware architecture. *Mirai* was found to support multiple hardware architectures, including *arm*, *mips*, *sparc* and *powerpc* (*step 4*).

The Scan/Loader server uses this information to login and instruct the vulnerable device to *tftp* or *wget* to the Scan/Loader server, download and execute the corresponding payload binary. Once executed, the first infected IoT device becomes part of the *Mirai* botnet and can communicate with the C&C server. The malware binary is removed and runs only in memory, to avoid detection (*step 5*). The botmaster can now issue attack commands, specifying parameters such as attack duration and target (*step 6*). The malware includes 10 DDoS attack types, including UDP flood (*udp*), Recursive DNS (*dns*), SYN packet flood (*syn*), ACK packet flood (*ack*), GRE flood (*gre ip*), which

[1] https://github.com/jgamblin/Mirai-Source-Code

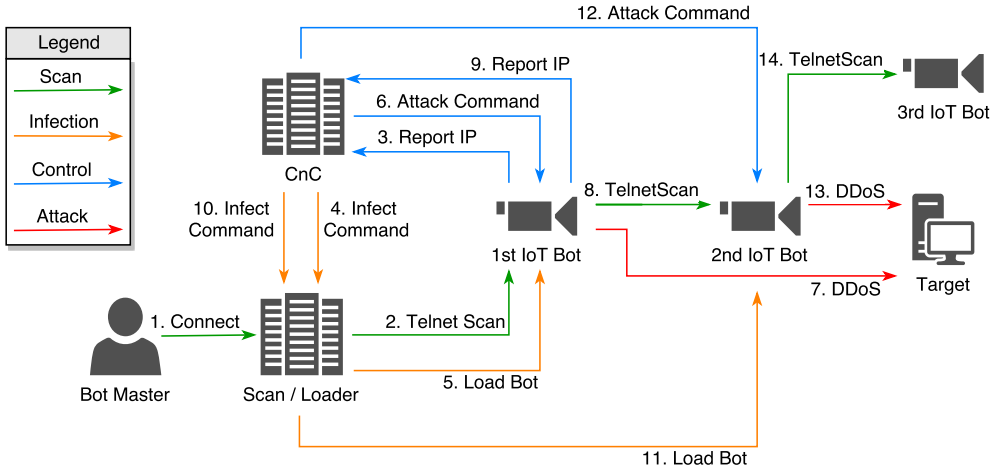


Figure 4.2: Botnet Infection and Proliferation

can be used to attack a target on the Internet (*step 7*). The first bot now attempts to repeat the infection process and propagate the botnet by scanning the Internet for additional vulnerable IoT devices with Telnet services and ports 23 or 2323 open (*step 8*). New vulnerable IoT device information is returned to the C&C server (*step 9*). A new infect command is issued to the Scan/Loader server (*step 10*). The appropriate hardware binary is loaded onto the newly discover vulnerable IoT device (*step 11*). The relevant attack command is issued from the C&C server (*step 12*). The attack is executed by the newly infected second bot, in conjunction with the first bot (*step 13*). Scanning for additional vulnerable IoT devices is repeated to further expand the botnet. (*step 14*).

4.2.4 Current Threat Detection: SNORT IDS

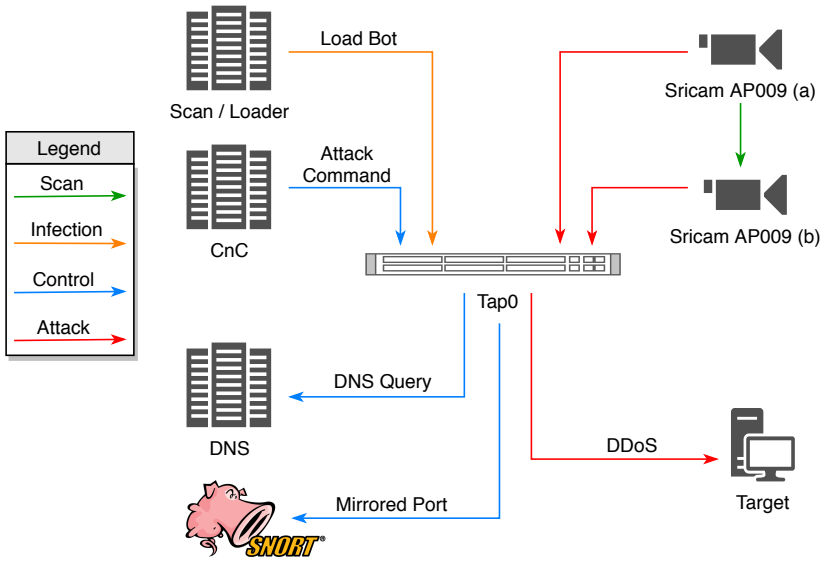


Figure 4.3: Snort IDS Experimental Setup

Chapter 2 explored the growing issue of security within the Internet of Things (IoT). It highlighted how household smart appliances often contain vulnerabilities [48] and are therefore targeted by hackers, threatening the security and privacy of families [82]. The chapter also demonstrated the need for security solutions to be developed, tailored specifically for the IoT, to enable users and organisations to better protect their smart devices [51]. In traditional networks, Intrusion Detection Systems (IDS) have proven to be powerful and versatile tools for network security management [179]. They have been successfully deployed to monitor the operations of corporate networks, generating system alerts when security violations are detected [180]. They are, however, not limited to the protection of complex networks or critical infrastructures, and can be configured to provide protection for smaller networks or single hosts. An aim of this study was to analyse the effectiveness of using an IDS to protect smart homes and devices within the IoT. Although a range of commercial and open-source IDS was found to exist, not all systems were found to be suitable for use in a smart home environment. Firstly, it was deemed unlikely that users in smart home environments would purchase an expensive commercial IDS; therefore, they were immediately omitted from consideration. Secondly, previous research had evaluated various open-source IDS for their effectiveness to detect threats and was also used to inform the selection of IDS to analyse. For example, in [181, 182] the authors compared **Snort**[1] and **Suricata**[2], two popular open source IDS, and found Suricata to scale better and handle larger volumes of traffic,

[1] https://www.snort.org/
[2] https://suricata-ids.org/

while returning comparable accuracy metrics as Snort. However, importantly Snort was found to have a lower system overhead in terms on CPU and RAM requirements. In the context of home environments it is also important to consider the requirements for computing resources, storage space and network bandwidth [183]. Here, the authors compared **Snort**, **Ourmon**[1], and **Samhain**[2] IDS and assessed them for CPU load, network bandwidth, and memory demand. Results confirmed the findings in [181, 182], and also found Snort to have a lowerer CPU load. Since scalability is unlikely to be a requirement in smart home environments, and home environments are likely to have limited resources on which to run an IDS, Snort was chosen as the IDS to be evaluated.

Snort was installed and added to the Botnet experimental setup (See Figure 4.3). A *Snort oinkcode* api key was generated, and *Pulled Pork* used to download the latest rule packages. The signature for the *Mirai* botnet[3] was also download and used to test the ability of *Snort* to detect infections and attacks. The *Mirai* botnet malware contains ten available attack vectors, which leverage infected IoT devices to engage in DDoS attacks against targets. To analyse *Snort* five attack vectors were chosen, including Acknowledgement *(ACK)* flood, Domain Name System *(DNS)* flood, User Datagram Protocol *(UDP)* flood, Generic Routing Encapsulation IP *(GREIP)* flood, and Synchronize *(SYN)* flood. Command and control messages between the C&C server and the infected IoT IP camera *(bot)* were also captured, as was normal traffic generated by the camera. Each attack was run for a period of 60 seconds, with *Snort* Log alerts summarised in Table 4.4.

4.2.5 Proposed Threat Detection: BLSTM-RNN IDS

Having analysed an existing method of threat detection, the study next explored if Deep Learning could be used to improve the detection of threats facing smart devices within the IoT. A new IDS model based on a Bidirectional Long Short Term Memory Recurrent Neural Network (BLSTM-RNN) was developed, and tested for its ability to detect the same attacks as used in Section 4.2.4. To test the model a dataset was generated from the experimental set-up described in Section 4.2.2, which consisted of *Mirai* botnet traffic such as *Scan*, *Infect*, *Control* and *Attack* traffic (described in Section 4.2.3), and also normal IoT IP Camera traffic. The captured dataset included features *No.*, *Time*, *Source*, *Destination*, *Protocol*, *Length*, and general information relating to the payload in the *Info* feature (See Table 4.1). All features were retained in the dataset, however, the model later removed features that provided limited scope for analysis such as *No.* and *Time*.

[1] http://ourmon.sourceforge.net/
[2] https://www.la-samhna.de/samhain/
[3] https://www.snort.org/rule_docs/1-40519

Finally, each packet was manually labelled as either normal or by its attack type (e.g. *ACK*, *UDP* etc).

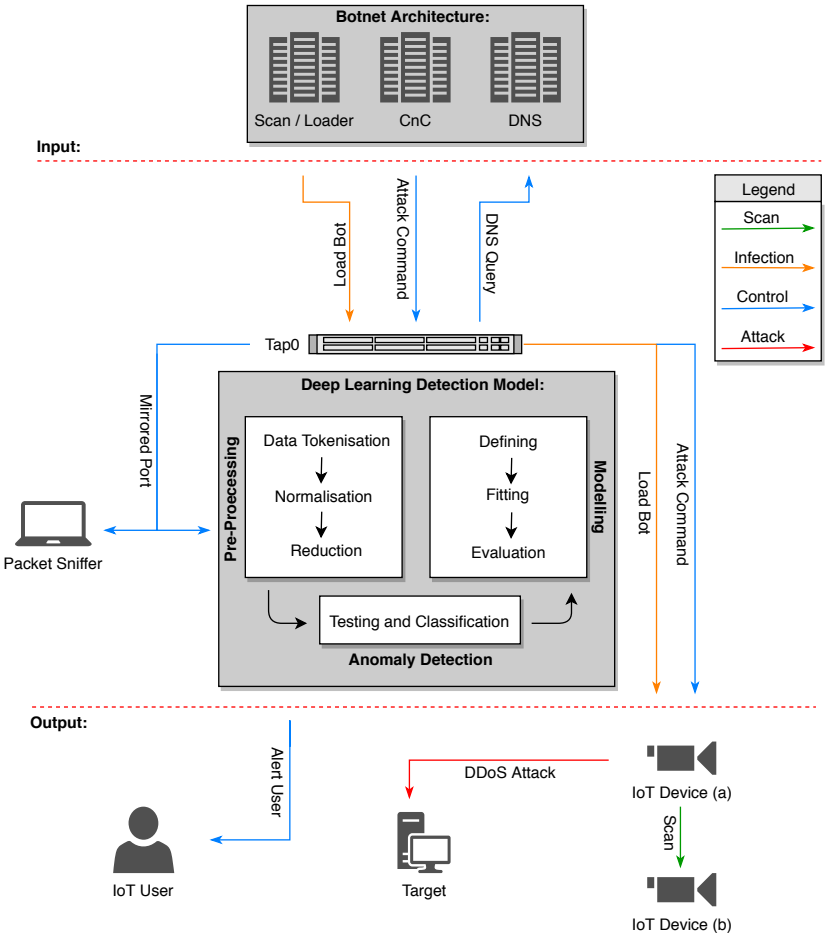


Figure 4.4: Botnet Architecture and Deep Learning Detection Model

Since a large portion of the captured information resided in the *Info* feature, as shown in Table 4.1 a model was required that could read and understand the text presented in this feature. As discussed in Section 2.1.3 an Artificial Neural Network(ANN) and more complex versions of Recurrent Neural Networks(RNN) such as Long Short Term Memory (LSTM) only work with numerical values. However, [184] demonstrated that a Deep Bidirectional Long Short Term Memory based RNN (BLSTM-RNN) can be used which provides promising results for text recognition. This potential was further demonstrated in [99, 100, 185] where a BLSTM-RNN was used in conjunction with Word Embedding, to map phrases and vocabulary to vectors or real numbers, and proved to be an effective method for modelling and predicting sequential text.

Table 4.1: Attack Packet Structure

Packet	Source	Destination	Pro	Len	Info
Normal	192.168.252.40	192.168.252.60	TCP	66	81 - 50451 [SYN, ACK] Seq=0 Ack=1 Win=5840 Len=0 MSS=1460 SACK_PERM=1 WS=2
Mirai	192.168.252.40	106.65.144.6	TCP	64	62002 - 23 [SYN] Seq=0 Win=57378 Len=0 [ETHERNET FRAME CHECK SEQUENCE INCORRECT]
UDP	192.168.252.40	192.168.252.50	UDP	554	55741 - 65170 Len=512
DNS	192.168.252.40	192.168.252.22	DNS	90	Standard query 0x0c9 A nnt1heibflkk.report.McDPhD.org

** No. and Time features Omitted for Brevity*

Table 4.2: ACK Packet Structure and Sequencing

Packet	Source	Destination	Pro	Size	Info
ACK	192.168.252.40	192.168.252.50	TCP	566	59693 - 41058 [ACK] Seq=1 Ack=1 Win=29597 Len=512
ACK	192.168.252.50	192.168.252.40	TCP	60	41058 - 59693 [ACK] Seq=1 Ack=1 Win=29597 Len=0
ACK	192.168.252.40	192.168.252.50	TCP	566	28029 - 45060 [ACK] Seq=1 Ack=1 Win=29597 Len=512
ACK	192.168.252.40	192.168.252.50	TCP	566	56493 - 64047 [ACK] Seq=1 Ack=1 Win=29597 Len=512

** No. and Time features Omitted for Brevity*

Motivated by the potential demonstrated by Wang et al. [185], and since the information captured in the *Info* feature of the dataset appeared to follow a sequence (See Table 4.2), a model was designed which used a BLSTM-RNN in conjunction with Word Embedding. This enabled string data to be converted into a format usable by the deep learning model. The approach taken was to first convert each letter into a tokenised and integer encoded format. Next, a dictionary of all tokenised words and their index within the *Info* feature was created and text replaced with its corresponding index number. In order to understand each attack type, it was important to maintain the sequence order of the indices, therefore an array of the indices was created. Since attacks are often closely coupled to the protocol used and the length of the captured packet, the *Protocol* and *Length* features also required to be included in the array. Word Embedding was again used to convert and create a dictionary of all tokenised protocols

and their index. These were then added, along with the *Length* feature, which was already an integer, to the array. Labels identifying each type of captured packets were mapped from string to integer ('norm': 0,'mirai':1,'udp': 2, 'dns':3, 'ack':4), and were also injected into the array. To simplify this process, the *Keras* library was used with a wrapper API around *Theano* and *Tensorflow*. The *Keras one_hot function* was used to convert strings into indices, form a 2-dimension list and create a dictionary at the same time. Finally, since deep neural networks require arrays to be of equal length, it was necessary to find the maximum length of a sentence within the *Info* feature and pad all the arrays with 0 to be equal to the maximum length of 25. After processing the dataset it was split into *training* and *test* datasets and reshaped into 3 dimensions, the format required for BLSTM-RNN layer (see Algorithm 1.)

To test the effectiveness of the deep learning approach to threat detection the model was tested against a series of attacks associated with the *Mirai* botnet. As shown in Algorithm 1 unit and Output layer with sigmoid activation were added to the model. The model was then compiled with the Mean Absolute Error (MAE) loss function and the Adam optimiser over total of 20 iterations, as shown in Table 4.3.

Table 4.3: BLSTM-RNN Model Parameters

Variables	**Values**
Activation	Sigmoid
Loss	Mean Absolute Error (mae)
Optimiser	Adam
BLSTM layer total units	20
Dense layer total unit	6
Epochs	20

Algorithm 1 BLSTM IoT Botnet Detection

```
1: procedure dataProcessing(attack dataset)
2: path ← attack dataset location
3: allFiles ← open pattern matched csv files in write mode
4: frame ← define two dimensional labeled data structure
5: unitToDrop ← 25%
6: repeat
7:     /*create concatenated dataset*/
8:     for i ← in allFiles do
9:         df ← read files
10:        list_ ← append(df) read files
11:    end for
12: until files concatenated into dataset
13: dataset ← concatenated (list_)
14: repeat
15:    /*Integer encode dataset*/
16:    for d ← in dataset.values do
17:        encoded_docs ← tokenise words
18:        dict ← create dictionary of encoded_docs
19:        array ← map indices of dict
20:        if array length != 25 then
21:            invoke-process ← pad array == 25
22:        end if
23:    end for
24: until Data tokenised and integer encoded
25: padded_docs ← array of tokenised and padded text
26: dataset.dropna ← split dataset based on unitToDrop
27: repeat
28:    /*Train and evaluate model*/
29:    model.compile ← (loss == mae, optimiser == adam)
30:    for i ← in epochs do
31:        reshape ← Training and Test to 3 Dimension
32:        model.evaluate ← Accuracy and Loss
33:    end for
34: until trainingDataset and testDataset are reshaped
35: Return Loss, ValLoss, Acc, ValAcc
```

Figure 4.4 shows the data flow and anomaly detection process. Data is transitioned through three distinct phases. The *Pre-processing* phase adjusts features to ensure data representation is suitable for the developed algorithm. Word Embedding tokenises the data, before normalisation and removal of packets with missing data is performed. In the *Modelling* phase the BLSTM-RNN algorithm is applied to the *training* data to define, fit and evaluate the detection model. Finally, in the *Anomaly Detection* phase the *test* data is tested to determine the effectiveness of the model in terms of accuracy and loss. Since the aim was to test the effectiveness of using a BLSTM-RNN with Word Embedding on sequential data, as demonstrated in [185], another model was required for comparison. A unidirectional LSTM-RNN was selected since it would only use past information for context, whereas the bidirectional LSTM-RNN would also utilise future contextual information [186]. Comparing the two models would allow us to ascertain if a bidirectional LSTM-RNN use of both past and future contextual information, would result in better accuracy or loss metrics for the captured threat dataset. Results of the comparative tests are presented in Table 4.6.

4.3 Results

4.3.1 SNORT IDS

The results for *Snort* analysis are presented. *Snort's* ability to detect malware and identify each of the five individual attacks was tested. Table 4.4 shows that *Snort* was able to detect the presence of mirai using the specific Snort signature ID (Sid)[1]. Sid *7748* "*MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt*" alerts were present in all attacks and appeared to relate to C&C messages used by mirai during the initial infection process and also when issuing subsequent attack commands. Sid *32655* "*stream5: TCP Small Segment Threshold Exceeded*" was also generated for each of the attacks. On first inspection, this alert would appear to be generated as a result of each of the DDoS flood attacks, however the generation of a TCP threshold alert would not be expected for the *UDP* based flood attack. Further research found that this alert plus Sids *32660, 32653* have previously been reported by some *Snort* users[2] who attributed it to the use of SSH/Telnet connections within the topology, not to the presence of an attack. It was therefore inconclusive whether these alerts were indicative of an attack taking place. The final alert *513* "*ICMP Test detected*" was only generated for three attacks *SYN, DNS, UDP*. It would appear Snort recognised this traffic as being anomalous, however was not able to attribute it specifically to the mirai botnet.

[1] https://tinyurl.com/mirai-virustotal
[2] https://seclists.org/snort/2012/q1/515

Table 4.4: Snort IDS Alerts

	Sid *	Description	Count
MIRAI	7748	MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt	2
	32655	stream5: TCP Small Segment Threshold Exceeded	39
ACK	7748	MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt	1
	32653	SSH: Protocol mismatch	12
	32655	stream5: TCP Small Segment Threshold Exceeded	21
SYN	513	ICMP Test detected	78
	7748	MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt	2
	32653	SSH: Protocol mismatch	36
	32655	stream5: TCP Small Segment Threshold Exceeded	45
DNS	513	ICMP Test detected	75
	7748	MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt	1
	32655	stream5: TCP Small Segment Threshold Exceeded	153
	32660	stream5: Reset outside window	2022
UDP	513	ICMP Test detected	72
	7748	MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt	2
	32655	stream5: TCP Small Segment Threshold Exceeded	60
GREIP	7748	MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt	2
	32655	stream5: TCP Small Segment Threshold Exceeded	54

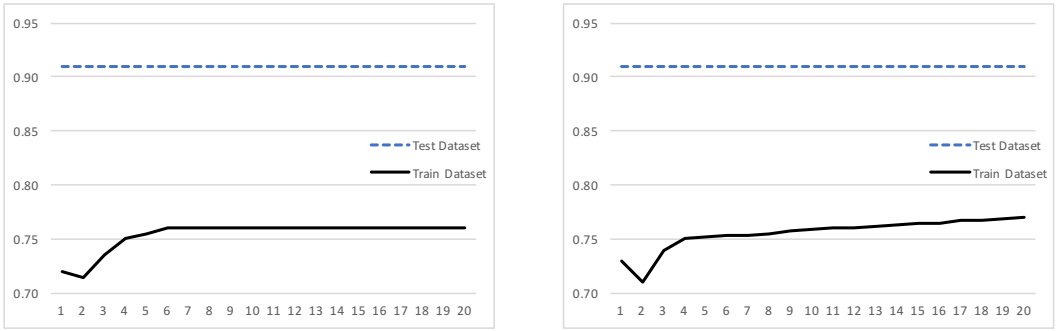
* ***Sid:*** *Snort Signature ID*

4.3.2 BLSTM-RNN IDS

Table 4.5: Captured Attack Samples

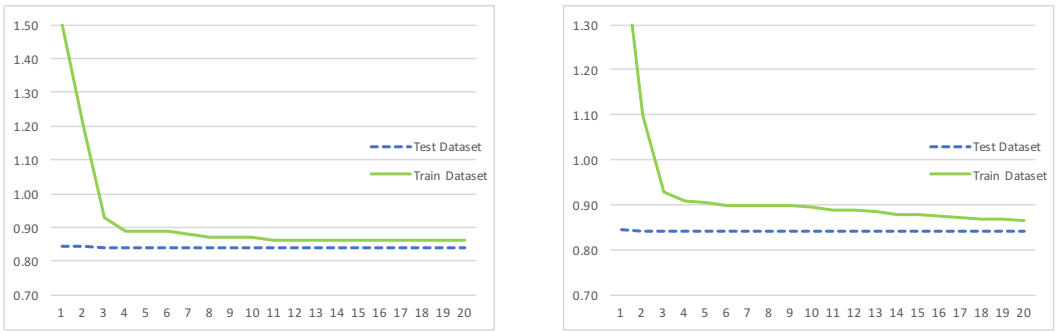
	Attack	Normal	Mirai	Cleaned
Mirai	0	598676	5102	595478
UDP	9380	590524	2576	601542
ACK	67444	588560	6372	632889
DNS	8706	598410	4408	602496

The results of the comparative test between the Bidirectional Long Short Term Memory Recurrent Neural Network (BLSTM-RNN) and the unidirectional LSTM-RNN are presented. To compare the deep learning detection models a series of four experiments were performed for each. Since unidirectional LSTM-RNN only preserve information from the past, the aim of the comparison was to ascertain if the use of a bidirectional LSTM-RNN, which is able to accumulate contextual information from both past and future, could return better accuracy or loss metrics for the captured dataset.



(**a**) LSTM Accuracy (**b**) BLSTM Accuracy

Figure 4.5: Accuracy Metrics for Detection Models



(**a**) LSTM Loss (**b**) BLSTM Loss

Figure 4.6: Loss Metrics for Detection Models

Table 4.6: Detection Accuracy and Loss

	Attacks (n)		Accuracy (%)		Loss	
	Train	**Validate**	**BLSTM**	**LSTM**	**BLSTM**	**LSTM**
Mirai	387060	208418	99.998992	99.571605	0.000809	0.027775
UDP	391002	210540	98.582144	98.521440	0.125630	0.125667
ACK	411384	221515	93.765198	93.765198	0.858700	0.858773
DNS	391622	210874	98.488289	98.488289	0.116453	0.116453
Multi-Vector (inc ACK)	419887	226094	91.951002	91.951002	0.841303	0.841381
Multi-Vector (no ACK)	395564	212996	97.521033	97.521033	0.115293	0.115293
Multi-Vector (three ACK)	468534	252289	92.243513	92.243513	0.161890	0.242358

For *Experiment 1* each attack type (See Table 4.5) was split between *train* and *validate* using a cross validation fold of 35%, presented to each model and trained over a total of 20 iterations. The mean accuracy and loss metrics for each attack were measured, and are presented in Table 4.6. As can be seen from the results, both models returned high accuracy and prediction for *mirai* (99%), *udp* (98%) and *dns* (98%) attack types. However, they returned less favourable results for *ack* (93%) attacks, despite this attack having the highest number of samples. Respective validation loss metrics *mirai* (0.000809), *udp* (0.125630), and *dns* (0.116453) were low, however, were again less favourable for the *ack* (0.858700) attacks.

Since multi-vector DDoS attacks were highlighted as being a growing issue in Section 2.1.2, *Experiment 2* consisted of *norm*, *mirai*, *udp*, *dns*, and *ack* captures being concatenated to form a multi-vector attack scenario. Results on row 5 of Table 4.6 show the impact of the *ack* attack on the overall detection accuracy (91%) and particularly loss metrics (0.841303). To validate this observation, *Experiment 3* consisted of *norm*, *mirai*, *udp*, and *dns* captures being concatenated to form a multi-vector attack scenario, without the *ack* attack. Results on row 6 of Table 4.6 show that once the *ack* attack was removed, overall detection accuracy and prediction (97%) of the model were very good. A final validation of this observation was conducted in *Experiment 4* which consisted of three *ack* attacks being performed during the same time frame, increasing the total sample size of *ack* attacks, in order to observe the variation in accuracy and prediction. Row 7 of Table 4.6 shows an increase in sample size, improved the overall validation accuracy to 92%, with BLSTM-RNN returning the better loss metric. Accuracy and loss metrics for both detection models are presented in Figures 4.5 to 4.6. It should be noted that the performance of the model has not been tested with new data due to a lack of associated IoT malware datasets at the time.

4.4 Discussion

This study was undertaken to investigate the effectiveness of current methods for detecting threats within the IoT and to present a new method of threat detection. The research in Chapter 2 demonstrated the growing issue of botnets within the IoT [52, 73] and how increasingly difficult it can be to detect these threats within smart home environments. It also demonstrated how easily malware can be mutated to create new variants of the original, making it difficult to maintain smart device security [68, 69, 70]. The *Mirai* malware was identified to be suitable for use in this study, since it predominately targeted consumer IoT devices, and the sourse code was readily available on github[1]. This allowed the experimental setup in Figure 4.1 to be established, which was then used to test an existing IDS, and the newly developed threat detection method presented in Section 4.2.5. *Snort* was chosen as a suitable IDS to analyse, and justification for its selection was provided in Section 4.2.4. From the results presented in Section 4.3.1, it was clear that *Snort* was able to detect the presence of the *Mirai* malware within the test environment, but was not able to detail each individual attack type. The *Snort* signature ID (Sid)[2] produced alert Sid *7748* "*MALWARE-CNC Unix.Trojan.Mirai variant post compromise echo loader attempt*" whenever the initial infection process was run, or whenever a new command was sent from the C&C server to an infected smart device (IoT Camera). It was not clear whether the remaining Sid alerts (513, 32653, 32655, 32660) could be attributed solely to the presence of *Mirai* since although they could be indicative of Telnet/SSH session use by *Mirai*, they have previously been reported by other *Snort* users[3] to be general alerts. Further investigation of the *Snort* rule for *Mirai* showed that the rule checks for specific string values used by the malware. This was evident from research in [10] where the authors presented the *Snort* rule as:

alert tcp $EXTERNAL_NET any <> $HOME_NET any (msg:"Possible Mirai infection"; content: "/bin/busybox MIRAI"; sid: 10003; rev:1;)

Here, the *Snort* rule uses the string */bin/busybox MIRAI* to trigger an alert, whenever a match is observed. Variations of this rule have also been found to monitor other strings used by *Mirai* such as *MIRAI: applet not found*, which was observed during the research in Section 4.2.3. While these rules appear to be effective for the original malware, they are likely to prove ineffective for future mutations. Indeed, from the mutated versions of *Mirai*, identified by avast [68] and presented in Table 2.1 six were found to use amended string values, which rendered the original *Snort* rules null and

[1] https://github.com/jgamblin/Mirai-Source-Code
[2] https://tinyurl.com/mirai-virustotal
[3] https://seclists.org/snort/2012/q1/515

void. The research did identify that new signature rules have been written to detect mutated variants of *Mirai*, such as satori:

alert tcp $EXTERNAL_NET any -> $TELNET_SERVERS 23 (msg:"BACKDOOR MISC Linux rootkit satori attempt"; flow:to_server,established; content:"satori"; reference:arachnids,516; classtype:attempted-admin; sid:216; rev:6;)[1]

However, writing new signature rules for every mutation of a malware would prove difficult to sustain long term. The research therefore demonstrated that the use of current signature-based intrusion detection, although accurate and effective at detecting known threats [82], can be ineffective at detecting new attacks or variants of known attacks, since a matching signature for these attacks is likely not known [187, 188]. The inability of existing signature-based IDS such as (*Snort*) to effectively detect new or variants of known IoT threats, was used as justification for developing the new threat detection method in Section 4.2.5.

Motivated by results in [184, 185] a threat detection model based on a Bidirectional Long Short Term Memory Recurrent Neural Network (BLSTM-RNN) was developed. In doing so, a method for modelling and predicting sequential text in the captured dataset was achieved. The BLSTM-RNN detection model was compared with a unidirectional LSTM-RNN to ascertain if the use of contextual information from both the past and future, could return better accuracy and loss metrics for the captured dataset. Results in Table 4.6 showed that both models returned high accuracy and prediction for *mirai* C&C commands, and *udp*, *dns* attack types. However, returned less favourable results for *ack* attacks, despite this attack having the highest number of samples. This was possibly due to the nature and complexity of information in the *info* feature, as seen in Table 4.2, where the sequence numbers in each *ack* packet changed. Despite this, a pattern can however be seen on rows one and two, where sequence numbers *(59693-41058, 41058-59693)* of contiguous packets were clearly linked, and packet size and Length were consistent. Unfortunately some packets appeared out of sync as seen in rows three and four, and possibly resulted in the detection model not recognising this pattern, contributing to the lower detection rate, and significantly higher loss metric. By contrast, although some captured packets in Table 4.1 appear to be equally complex, the information in the *info* feature for each packet type, remained largely the same, possibly aiding better detection.

The BLSTM-RNN model was assessed for its effectiveness to detect multi-vector DDoS attacks and results showed that overall detection accuracy and prediction was still generally very good. However, since the model did not perform as well for *ack* attacks,

[1] https://github.com/eldondev/Snort/blob/master/rules/backdoor.rules

when these were present in a multi-vector attack overall accuracy was reduced (91%). Interestingly Figures 4.5 to 4.6 show the accuracy and loss metrics for the LSTM-RNN appeared to level after 10 iterations, however, continued to improve with each additional iteration for the BLSTM-RNN. Therefore, the total sample size of *ack* attacks was increased, in order to observe the variation in accuracy and prediction. Results showed that the increase in sample size, resulted in improved accuracy (97%), and better loss metrics in the BLSTM-RNN model (0.161890). The results of this study showed that the bidirectional nature of the BLSTM-RNN, and its use of contextual information from the past and future, coupled with a larger datatset made it a better progressive model over time. Training the model with a larger dataset could result in further improvements in accuracy and loss. This study set out to answer the question "*Can current security methods detect the presence of threats within consumer IoT networks ?*". It is clear from the results that although the existing IDS (*Snort*) was able to detect the presence of threats (*Mirai*) used in the study, mutated versions of the malware could prove more difficult to detect, rendering existing signatures ineffective. The proposed threat detection model (BLSTM-RNN) demonstrated the potential for deep learning to be applied to threat detection in consumer IoT networks. The study showed that once trained with previous attack data the deep learning IDS model could accurately predict future threats facing consumer IoT network.

4.5 Publication of Dataset

During the undertaking of this study a labelled IoT botnet dataset was created, containing both normal traffic and attacks [17] from the *Mirai* malware. The data set spans five days and incorporates a total of 37 *wireshark pcap* files, and corresponding labelled *csv* files. The generated *mirai* botnet dataset has been made public, been cited extensively[1], and used for comparative studies [14, 15]. The dataset is available upon request[2].

[1] https://tinyurl.com/rfayhz6
[2] https://tinyurl.com/ResourcesCMcD

4.6 Conclusions

This chapter examined the ability of a current security method (*Snort*) to detect threats within consumer IoT networks. In Section 2.1.2 botnets were identified to be a particular problem facing the IoT, therefore the *Mirai* malware was chosen as a suitable threat for use in this study, since it predominately targeted consumer IoT devices. The results of this study showed the IDS to be accurate and effective at detecting the *Mirai* malware, however also demonstrated it could be ineffective at detecting new variants of the malware. Finally, a new threat detection method was presented based on a BLSTM-RNN. Once trained, the model accurately predicted future threats from the *Mirai* malware and demonstrated the potential for deep learning to be used for threat detection. In the next chapter, awareness and perception of threats within consumer IoT networks will be examined and the results of a cross-sectional study, which examined if users are able to detect the presence of botnet activity, will be presented.

Chapter 5

Situational Awareness of Threats in Consumer IoT Networks

In Chapter 4 the effectiveness of current threat detection methods in consumer IoT networks was examined. A recent and prominent example (*Mirai botnet*) of threat facing consumers IoT networks was analysed, to better understand how insecure smart devices can be exploited and leveraged to perform attacks on the Internet. A new application of deep learning for threat detection was also proposed and clearly demonstrated its effectiveness in detecting botnet activity within consumer IoT networks. This chapter examines the awareness and perception of threats within consumer IoT networks. Previous work [189, 190, 191] had suggested that demographic characteristics may have an effect on users awareness of threats. Results of a cross-sectional study are presented which examined this phenomenon and also if users are able to detect the presence of botnet activity.

5.1 Introduction

The focus of this chapter is attempting to answer sub research question SQ2: "*Can users visually detect the presence of threats within consumer IoT networks ?*". Firstly, a quantitative approach is used to examine how users value and perceive security and privacy in smart devices found within the IoT. Next, user requirements from IoT devices are analysed and the importance placed upon security and privacy investigated. Secondly, the ability of users to detect threats is assessed, in the context of demographic characteristics, namely technical knowledge and age. This twin-pronged approach to analysis is carried out to examine the impact of botnets within the IoT, in the context of how they are perceived within consumer environments.

5.2 Methodology

5.2.1 Experimental Variables

The study presented in this chapter used an online survey, which was split into two sections. The first section collected information relating to user awareness and requirements of security and privacy, and is discussed in Section 5.2.2. The second section measured participants ability to identify when a smart device had been infected, and was being used to perform attacks on the Internet. Detection accuracy was used as the dependent variable and data collected was dichotomous and nominal. In [3], the author suggests a lack of technical knowledge and ability to explore network communication, results in little or no awareness of security issues. An aim of this study was to test this assumption and discover if there was a relationship between the dependent variable and participant's technical knowledge which was used as the independent variable and collected ordinal data. A study in [190] also suggested age may have impact on security awareness, where they found participants aged over 55, although heavy users of "gadgets" (smart devices), overwhelmingly failed to recognise threats, and neglected to protect their connected devices. A second aim of this study was to test this assumption, and discover if age has an impact on security. Age was, therefore, used as the second independent variable, and used to determine if a relationship existed between the dependent variable and a participant's age. Age was categorised into ranges, and was collected as categorical data. The following hypothesis is derived from the desire to test these assumptions:

Hypothesis A: Demographic characteristics have an effect on the accuracy of detecting threats in consumer networks

The null hypothesis for a chi-square independence test states that two categorical variables are independent in a population [170]. To test the assumption that no association existed between the dependent and independent variables it was hypothesised:

> **H_1:** There is no association between **detection accuracy** and **technical knowledge** when detecting threats in consumer networks.
>
> **H_2:** There is no association between **detection accuracy** and **age** when detecting threats in consumer networks.

5.2.2 Study Design

An online survey instrument was produced and used to assess how users value and perceive security and privacy in smart devices found within the IoT. The online survey was created using the guidelines specified in Section 3.3.3, and was split into two

sections, each comprised of 17 questions in total. Section one collected information relating to user awareness and requirements of security and privacy in smart devices contained within the IoT. Section two evaluated user's ability to identify when a smart device had been infected, and was being used to perform attacks on the Internet. The secure sand-boxed environment created in Section 4.2.2 was used to perform DDoS attacks against a smart camera, commonly found and exploited within the IoT [192]. Three DDoS attack scenarios were performed and recorded, including a DNS flood attack, Synchronise (SYN) flood attack, and a Generic Routing Encapsulation over IP (GREIP) flood attack. Normal traffic was also generated from an uninfected camera, and was used for comparison. Participants were presented with the four recorded scenarios, and asked to identify if an attack had taken place, specifying the time when they think it occurred.

5.2.3 Participants

The aim of the study was to assess how users value and perceive security and privacy in smart devices found within the IoT. The study analysed user requirements from IoT devices, and the importance placed upon security and privacy. Convenience sampling was employed, with participants selected due to their convenient accessibility, and proximity to the author. Participants were recruited between October and November 2018 from the local university population, and also through LinkedIn and Social Media. This approach enabled a wide range of views to be collected and avoid oversampling of a specific demographic. Participants provided informed consent by reading the study agreement on the first page, before indicating their consent to participate when clicking to proceed to the next page of the survey. A total of one hundred ninety two participants started the study, however, thirty four did not complete resulting in an attrition rate of 17.7%. The threat of attrition in research to internal and external validity is an important issue [193]. Attrition rates of 30-40% are indicative of "fatal" flaws within a study, while below 20% is acceptable [194]. Therefore, the results of this study are believed to have validity, since the attrition rate was below the acceptable level. Table 5.1 presents the participant demographics where 17 (11%) participants were aged [under 18], 52 (33%) aged [18-24], 54 (34%) aged [25-39], 29 (18%) aged [40-59], and 6 (4%) aged [60+]. When asked to indicate their level of technical knowledge 23 (15%) self-identified as [Novice], 70 (44%) as [Intermediate], 57 (36%) as [Advanced], and 8 (5%) as [Expert]. Participants had a varying range of computing experience with 42 (27%) currently working within a computing related environment, 86 (54%) currently studying, and 30 (19%) not currently studying or working within a computing related environment.

Table 5.1: Participant Demographic

Age		%	**Ability**		%
< 18	17	(11)	**Novice**	23	(15)
18-24	52	(33)	**Intermediate**	70	(44)
25-39	54	(34)	**Advanced**	57	(36)
40-59	29	(18)	**Expert**	8	(5)
60+	6	(4)			

n = 158

5.3 Results

5.3.1 Section One Results

In Section one of the online survey participants were asked a series of questions relating to their awareness and perception of security and privacy considerations and requirements in IoT devices. Firstly, as shown in Figure 5.1a, participants were asked if they owned any IoT devices; 70 (44%) responded that they did not own any IoT devices, 57 (36%) owned one device, with *Amazon Echo* being the most popular with 47 (30%) respondents. 31 (20%) respondents indicated they owned two or more IoT devices.

(**a**) Devices owned by participants

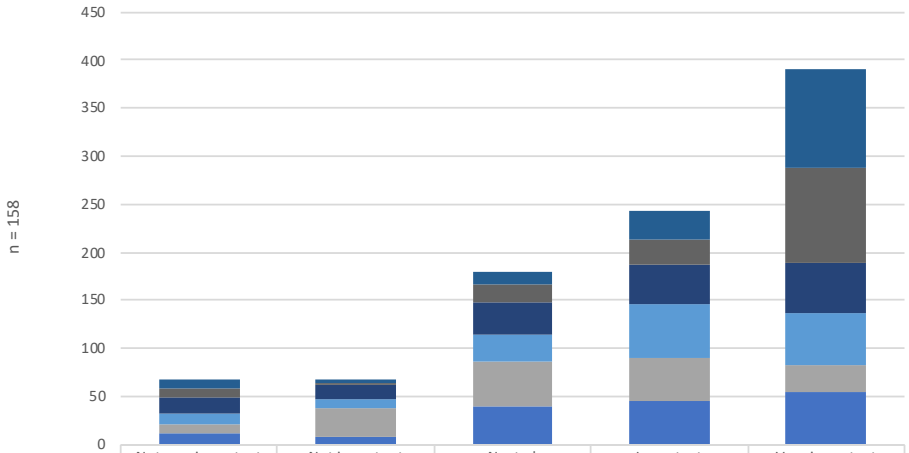
(**b**) Level of concern

Figure 5.1: Exposure to IoT devices and level of security concern.

To measure perception and importance placed on security and privacy, respondents were asked to rate the importance of various features related to IoT devices. As shown in Figure 5.2 security 102 (65%) and privacy 100 (63%) were clearly considered very important features by a large percentage of the population. However, interestingly when asked to rank the features in order of priority, cost was ranked higher than both security and privacy by the largest percentage of respondents 53 (34%) (see Figure 5.3). Although compatibility and ease of setup were considered very important features as

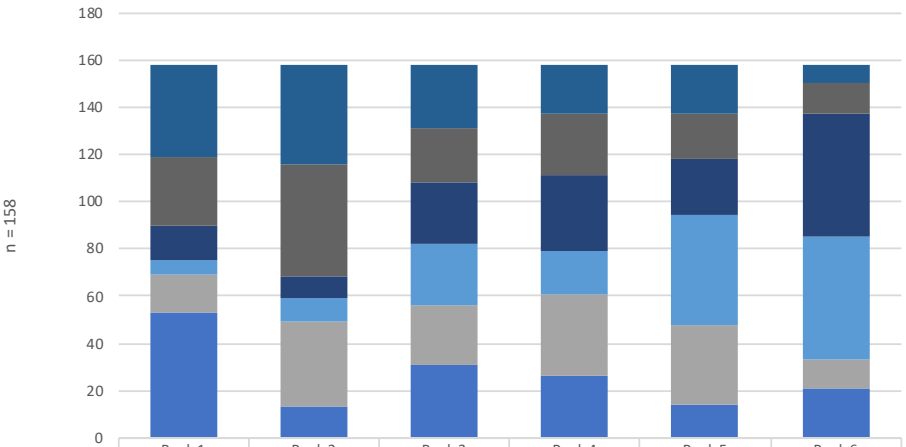
shown in Figure 5.2, again when asked to rank features in order of priority they were very clearly ranked less important (see Figure 5.3).

To assess whether respondents ranked security and privacy highly in theory, but not in practice, respondents were asked how concerned they would be if a smart device they owned was infected with a virus, but was still functioning as expected. Figure 5.1b shows that over three quarters of respondents 91 (58%) and 41 (26%) respectively said they would be very concerned or concerned.



	Not very important	Not important	Neutral	Important	Very important
Security	9	4	14	29	102
Privacy	10	2	19	27	100
Compatibility	16	15	33	42	52
Ease of Setup	12	9	28	55	54
Ease of Use	9	30	47	45	27
Cost	11	8	39	45	55

Figure 5.2: IoT device feature importance.



	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
Security	39	42	27	21	21	8
Privacy	29	48	23	26	19	13
Compatibility	15	9	26	32	24	52
Ease of Setup	6	10	26	18	46	52
Ease of Use	16	36	25	35	34	12
Cost	53	13	31	26	14	21

Figure 5.3: IoT device feature rank.

5.3.2 Section Two Results

In section two of the online survey respondents were presented with the four recorded scenarios in Section 5.2.2. Scenario *A* and *B* were presented as live video feeds from an IoT IP camera. In scenario *A* the camera was not infected, and no attack was performed. In scenario *B* the camera was infected and performed a *DNS* flood attack against a victim device in the sandboxed environment (see Figure 5.4). Scenario *C* and *D* were presented as recorded outputs from a popular packet capture tool *(wireshark)*. In scenario *C* the camera was infected and performed a *SYN* flood attack against a victim device in the sandboxed environment. In scenario *D* the camera was infected and performed a *GREIP* flood attack (see Figure 5.5).

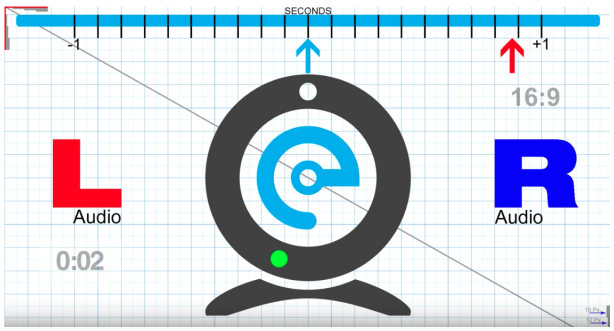
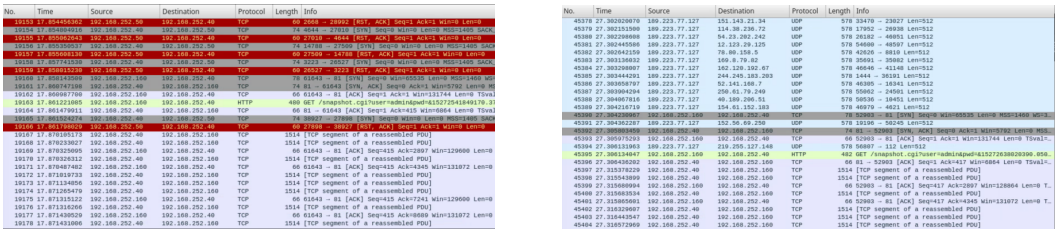


Figure 5.4: IoT IP Camera Video Feed (Scenario A and B).



(**a**) SYN Flood Attack (Scenario C)

(**b**) GREIP Flood Attack (Scenario D)

Figure 5.5: IoT Camera Video Feed (Scenario C and D).

Results for the four scenarios are presented in Tables 5.2–5.5 with time periods highlighted to indicate when the associated attack took place. In scenario *A* 97 (61%) of respondents indicated they could not tell if the IoT IP camera was infected and an attack took place (see Table 5.2). Respondent responses were consistent across all time periods. In scenario *B* an attack took place during time period [31–50 s]. Again, most respondents 94 (59%) indicated they could not tell if the IoT IP camera was infected and an attack took place. Respondent responses were again consistent across all time periods. When asked how easy it was to identify when the IoT IP camera was infected, 32 (38%) indicated very difficult, 25 (29%) difficult (see Figure 5.6a), indicating that it was not easy to detect if the device was infected from the presented live video feed shown in Figure 5.4. This was consistent with the authors own observations that during the infection process and attacks, the camera did not display any adverse symptoms of infection, and continued to function as expected. Remote access to the device was still possible, and performance did not appear to be degraded. Live video streaming continued to be as responsiveness as prior to the attacks. Therefore, without any clear signs of an infection it was confirmed that detection or awareness of botnet activity proved very difficult within consumer networks.

Table 5.2: Scenario A: Detection rate (no attack).

	0–10 s	**11–20 s**	**21–30 s**	**31–40 s**	**41-50 s**	**51–60 s**	**Dont Know**
	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**
Yes	22 (14)	24 (15)	26 (16)	23 (15)	24 (15)	13 (8)	97 (61)
No	136 (86)	134 (85)	132 (84)	135 (85)	134 (85)	145 (92)	61 (39)

Table 5.3: Scenario B: Detection rate (dns attack).

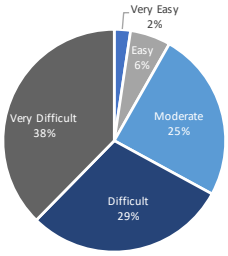
	0–10 s	**11–20 s**	**21–30 s**	**31–40 s**	**41–50 s**	**51–60 s**	**Dont Know**
	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**
Yes	35 (22)	26 (16)	33 (21)	29 (18)	30 (19)	23 (15)	94 (59)
No	123 (78)	132 (84)	125 (79)	129 (82)	128 (81)	135 (85)	64 (41)

Table 5.4: Scenario C: Detection rate (syn attack).

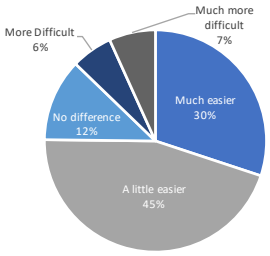
	0–10 s	**11–20 s**	**21–30 s**	**31–40 s**	**41–50 s**	**51–60 s**	**Dont Know**
	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**
Yes	36 (23)	76 (48)	92 (58)	53 (34)	30 (19)	27 (17)	38 (24)
No	122 (77)	82 (52)	66 (42)	105 (66)	128 (81)	131 (83)	120 (76)

Table 5.5: Scenario D: Detection rate (greip attack).

	0–10 s	**11–20 s**	**21–30 s**	**31–40 s**	**41–50 s**	**51–60 s**	**Dont Know**
	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**	***n* (%)**
Yes	44 (28)	34 (22)	39 (25)	47 (30)	51 (32)	30 (19)	65 (41)
No	114 (72)	124 (78)	119 (75)	111 (70)	107 (68)	128 (81)	93 (59)



(**a**) No Attack and DNS Attack



(**b**) SYN and GREIP Attack

Figure 5.6: User perception of detection difficulty.

In scenario *C* an attack took place during time period [11–30 s] and participants were shown the recorded *wireshark* output (see Figure 5.5a). The use of the packet capture tool significantly improved detection of the infected IoT IP camera with 120 (76%) of

respondents now indicating they knew when an attack took place. Results in Table 5.4 confirm this, with 76 (48%) [11–20 s] and 92 (58%) [21–30 s] correctly identifying the time period when the attack took place. In scenario *D* an attack took place during time period [21–40 s] and participants were shown the recorded *wireshark* output (see Figure 5.5b). In this scenario the packet capture tool did not appear to improve detection, as results presented in Table 5.5 show respondent responses were varied across all time periods. The number of respondents who indicated they knew when the attack took place dropped with 93 (59%) of respondents now indicating they knew when an attack took place.

Bivariate analysis was employed to cross tabulate between variables and look for possible associations. To determine if an association existed between a participants level of technical knowledge and their ability to detect an attack, a cross tabulation between these variables was undertaken. Frequency distributions were calculated and are presented in Tables 5.6–5.9.To check for relationships between the two categorical variables Chi-square tests were performed to check for independence [170] and examine the association between technical knowledge level and the ability to detect attacks. The data met the assumptions of having degrees of freedom greater than one (more than one group being compared), randomised samples and independent observations [171].

Table 5.6: Scenario A: Accuracy by Knowledge Level (no attack).

Knowledge	**Novice**	**Intermediate**	**Advanced**	**Expert**
	n (%)	***n (%)***	***n (%)***	***n (%)***
Yes	19 (83)	39 (56)	30 (53)	7 (88)
No	4 (17)	31 (44)	27 (47)	1 (12)
n = *158*, p = *.026*				

Table 5.7: Scenario B: Accuracy by Knowledge Level (dns attack).

Knowledge	**Novice**	**Intermediate**	**Advanced**	**Expert**
	n (%)	***n (%)***	***n (%)***	***n (%)***
Yes	0 (0)	15 (21)	15 (26)	1 (12)
No	23 (100)	55 (79)	42 (74)	7 (88)
n = *158*, p = *.028*				

Table 5.8: Scenario C: Accuracy by Knowledge Level (syn attack).

Knowledge	**Novice**	**Intermediate**	**Advanced**	**Expert**
	n (%)	***n (%)***	***n (%)***	***n (%)***
Yes	11 (48)	41 (59)	37 (65)	6 (75)
No	12 (52)	29 (41)	20 (35)	2 (25)
n = *158*, p = *.584*				

Table 5.9: Scenario D: Accuracy by Knowledge Level (greip attack).

Knowledge	**Novice**	**Intermediate**	**Advanced**	**Expert**
	n (%)	***n (%)***	***n (%)***	***n (%)***
Yes	2 (9)	28 (40)	23 (40)	5 (62)
No	21 (91)	42 (60)	34 (60)	3 (38)
n = *158*, p = *.013*				

Table 5.10: Accuracy within knowledge level (all scenarios).

Knowledge	**Scenario A (no)**	**Scenario B (dns)**	**Scenario C (syn)**	**Scenario D (greip)**	**Total**
	n (%)	***n (%)***	***n (%)***	***n (%)***	***n (%)***
Novice	19 (83)	0 (0)	11 (48)	2 (9)	32 (35)
Intermediate	39 (56)	15 (21)	41 (59)	28 (40)	123 (44)
Advanced	30 (53)	15 (26)	37 (65)	23 (40)	105 (46)
Expert	7 (88)	1 (12)	6 (75)	5 (62)	19 (59)

(%) *percentage of accurate detections for each scenario*

Scenario *A* was used as a control, however since users were not informed of this, attempts were still made and are presented in Table 5.6. Since an attack did not exist, if a user selected 'No' against each time scale and indicated they 'Don't Know' if the device was infected, this was used as evidence of a correct detection. It was not possible to analyze whether an association existed between knowledge level and the ability to detect an infected device, for this scenario. The results were, however, found to be statistically significant, χ^2 (3, n =158) = 9.253, p = .026. In scenario *B*, a dns attack

was performed, but a clear association between level of knowledge and ability to detect an infected device, was not evident. Results were again found to be significant, χ^2 (3, n =158) = 9.094, p = .028. In scenario *C*, a *syn* attack was performed, and participants were shown output from a packet capture tool *wireshark* (see Figure 5.5a). Detection rates across all knowledge levels increased substantially and a positive association was evident between knowledge level and a user's ability to detect an attack. Results were, however, not found to be statistically significant, χ^2 (3, n =158) = 1.944, p = .584. Finally, in scenario *D*, a *greip* attack was performed, and participants were shown output from a packet capture tool *wireshark* (see Figure 5.5b). Again there appeared to be a general positive association between the variables, however frequency distributions in Table 5.9 indicated the association was not as clear compared to scenario *C*. Results were found to be significant, χ^2 (3, n =158) = 10.711, p = .013. Accuracy levels for each technical knowledge level are summarised in Table 5.10.

Bivariate analysis was also employed to determine if an association existed between a participants age and their ability to detect an attack. To check for relationships between the two categorical variables Chi-square tests were performed to check for independence [170] and examine the association between age and the ability to detect attacks. Frequency distributions were calculated and are presented in Tables 5.11–5.14. Accuracy levels for each age group are summarised in Table 5.15 and did not indicate an association existed between age and ability to detect an attack. Results were not found to be significant across scenarios *A*-*D* (p = .268, .120, .190, .127 respectively).

Table 5.11: Scenario A: Accuracy by Age (no attack).

Age	**< 18**	**18-24**	**25-39**	**40-59**	**60+**
	n (%)	***n (%)***	***n (%)***	***n (%)***	***n (%)***
Yes	7 (41)	28 (54)	33 (61)	23 (79)	4 (67)
No	10 (59)	24 (46)	21 (39)	6 (21)	2 (33)
n = *158,* p = *.268*					

Table 5.12: Scenario B: Accuracy by Age (dns attack).

Age	**< 18**	**18-24**	**25-39**	**40-59**	**60+**
	n (%)	***n (%)***	***n (%)***	***n (%)***	***n (%)***
Yes	5 (29)	13 (25)	11 (20)	1 (3)	2 (33)
No	12 (71)	39 (75)	43 (80)	28 (97)	4 (67)
n = *158*, p = *.120*					

Table 5.13: Scenario C: Accuracy by Age (syn attack).

Age	**< 18**	**18-24**	**25-39**	**40-59**	**60+**
	n (%)	***n (%)***	***n (%)***	***n (%)***	***n (%)***
Yes	13 (76)	34 (65)	31 (57)	12 (41)	4 (67)
No	4 (24)	18 (35)	23 (43)	17 (59)	2 (33)
n = *158*, p = *.190*					

Table 5.14: Scenario D: Accuracy by Age (greip attack).

Age	**< 18**	**18-24**	**25-39**	**40-59**	**60+**
	n (%)	***n (%)***	***n (%)***	***n (%)***	***n (%)***
Yes	10 (59)	20 (38)	17 (31)	6 (21)	1 (17)
No	7 (41)	32 (62)	37 (69)	23 (79)	5 (83)
n = *158*, p = *.127*					

Table 5.15: Detection accuracy within age level (all scenarios).

Age	Scenario A (no)	Scenario B (dns)	Scenario C (syn)	Scenario D (greip)	Total
	n (%)	*n (%)*	*n (%)*	*n (%)*	*n (%)*
< 18	7 (41)	5 (29)	13 (76)	10 (59)	35 (51)
18 - 24	28 (54)	13 (25)	34 (65)	20 (38)	95 (46)
25 - 39	33 (61)	11 (20)	31 (57)	17 (31)	92 (43)
40 - 59	23 (79)	1 (3)	12 (41)	6 (21)	42 (36)
60 +	4 (67)	2 (33)	4 (67)	1 (17)	11 (46)

% shows the percentage of accurate detections for each scenario

5.4 Discussion

This study was undertaken to investigate user awareness and perception of security and privacy within the IoT. Since the *Mirai* malware predominately targeted consumer IoT devices, it was chosen for use in the experimental setup. In the process of building the experimental setup shown in Figure 4.1 it became clear how easily botnet malware can spread, and new variants and mutations of existing botnets appear on the Internet. Indeed, this is evident in [69, 70] where *satori*, *masuta*, *wicked*, and *JenX* are presented as new variants of the original *Mirai* botnet. Sharing the original basecode with *Mirai*, these new variants are enhanced to allow direct control of compromised devices, making other malicious actions possible, including running trojan viruses, redirecting traffic for *man-in-the-middle* attacks, and delivering other viruses to devices on the network by proxy. The last point being particularly concerning, since devices which were not originally vulnerable, could now be infected. In this study, 56% of respondents indicated they owned an IoT device, with 20% owning one more device. The study found the *Amazon Echo* to be the most popular IoT device (30%); however, many IoT devices leveraged by the above botnets, such as smart lightbulbs (16%) and IP cameras (8%), were also popular. Despite IP cameras only accounting for 8% of devices, if they could be leveraged and used as a proxy to infect other devices in home networks, the potential impact from IoT botnets, could be significantly greater than already experienced. Clearly, early detection and mitigation of such attacks is vital.

The study first explored participants attitudes towards security and privacy in the IoT. To assess whether respondents ranked security and privacy highly in theory, but not in practice, respondents were asked how concerned they would be if a smart device

they owned was infected with a virus, but was still functioning as expected. In asking this question, the aim was to assess whether the phenomenon known as the *Privacy Paradox* was evident in the context of attitudes towards IoT devices. The privacy paradox has been well documented in papers such as [195, 196], and although mainly in the context of online security, demonstrates that user attitudes towards security and privacy, often differ from the actions they take or decisions they make. Indeed, this has been highlighted in studies such as [19, 197]. In this study, Figure 5.1b shows that given a scenario where a device was infected with malware, but still functioning normally, over three quarters of respondents indicated that they would still be very concerned. When asked to rate the importance of various features related to IoT devices (as shown in Figure 5.2), security 102 (65%) and privacy 100 (63%) were clearly considered very important features. However, interestingly when asked to rank the features in order of priority, cost was ranked higher than both security and privacy by the largest percentage of respondents 53 (34%) (see Figure 5.3). In [198] it is suggested, while many users show theoretical interest in their privacy and maintain a positive attitude towards privacy-protection behaviour, this rarely translates into actual protective behaviour. The results in this study could confirm this, and suggest a possible dichotomy between privacy attitudes and actual behaviour, during procurement of IoT devices.

The study next explored the particpants ability to detect threats in consumer IoT networks. In doing so, a sandboxed botnet environment was used to infect an IoT IP camera, and leverage it to perform four attacks against a target. Respondents of the online survey were presented with video recordings of the four recorded attack scenarios, and their situational awareness and ability to detect infections recorded. Situational Awareness (SA) was defined in Chapter 2 as "the state of being aware of circumstances that exist around us, especially those that are particularly relevant to us and which we are interested about" [199]. Applied in a cyber context the author further presents an adapted SA model composed of four levels where perception, deals with evidence gathering of situations in the network. Comprehension refers to the analysis of evidence to deduce threat level, type and associated risk. Projection deals with predictive measures to address future incidents, and resolution deals with controls to repair, recover and resolve network situations [122]. This study evaluates the first of these levels (perception), and clearly demonstrates the difficulty users face in detecting threats found in IoT consumer networks. In scenario A and B users were presented with video recordings as shown in Figure 5.4. During the infection process and attacks, participants indicated that the camera did not display any adverse symptoms of infection, and continued to function as expected. This was evident from the results in Tables 5.2 and 5.3 where 61% and 59% of respondents reported not being able to detect any unusual activity in the video. Comments from respondents included:

"There wasn't any clear evidence" (Advanced Participant)

"I could not tell at all if the camera was infected" (Intermediate Participant)

In [122] the author suggests that perception in the context of Cyber SA also refers to knowledge of the elements in the network, and awareness of alerts such as those reported by intrusion detection systems, firewall logs, and scan reports. However, while this is certainly true of security analysts, this information is unlikely to be available in consumer networks, therefore may not be a contributing factor in achieving SA in consumer networks. In these environments the user would only have information displayed by the IoT device, in the case of scenario *A* and *B* in the study that would be the live video feed. Since there were no adverse symptoms of infection, and the IP camera continued to function as expected, it is understandable that 32 (38%) indicated it was very difficult, and 25 (29%) difficult, to detect the device was infected from the presented live video feed.

In scenario *C* and *D* users were presented with recorded outputs from a popular packet capture tool *(wireshark)* as shown in Figure 5.5. The use of the packet capture tool significantly improved detection in scenario *C* with 120 (76%) of respondents now indicating they knew when an attack took place. Results in Table 5.4 confirm this, with 76 (48%) [11–20 s] and 92 (58%) [21–30 s] correctly identifying the time period when the attack took place. However, in scenario *D* the packet capture tool did not appear to improve detection, as results presented in Table 5.5 show respondent responses were varied across all time periods. The number of respondents who indicated they knew when the attack took place also dropped to 93 (59%).

In [3], the author presents the need for greater online awareness and protection for NEUs. The author undertook a study to establish the views of NEUs on personal cyber security and suggests a lack of technical knowledge and ability to explore network communication, results in little or no awareness of security issues. Previous studies such as [191] have also demonstrated relationships between the technical ability of a user, and the ability to be perceive and be aware of risks. To test this assumption a hypothesis was defined in section 5.2.1 as:

> **H_1:** There is no association between **detection accuracy** and **technical knowledge** when detecting threats in consumer networks.

Figure 5.7 shows the accuracy of detection for each scenario across the four knowledge levels. When accuracy rates for the four scenarios are combined to give a *Total* accuracy rating a positive association between the two variables did appear to be evident.

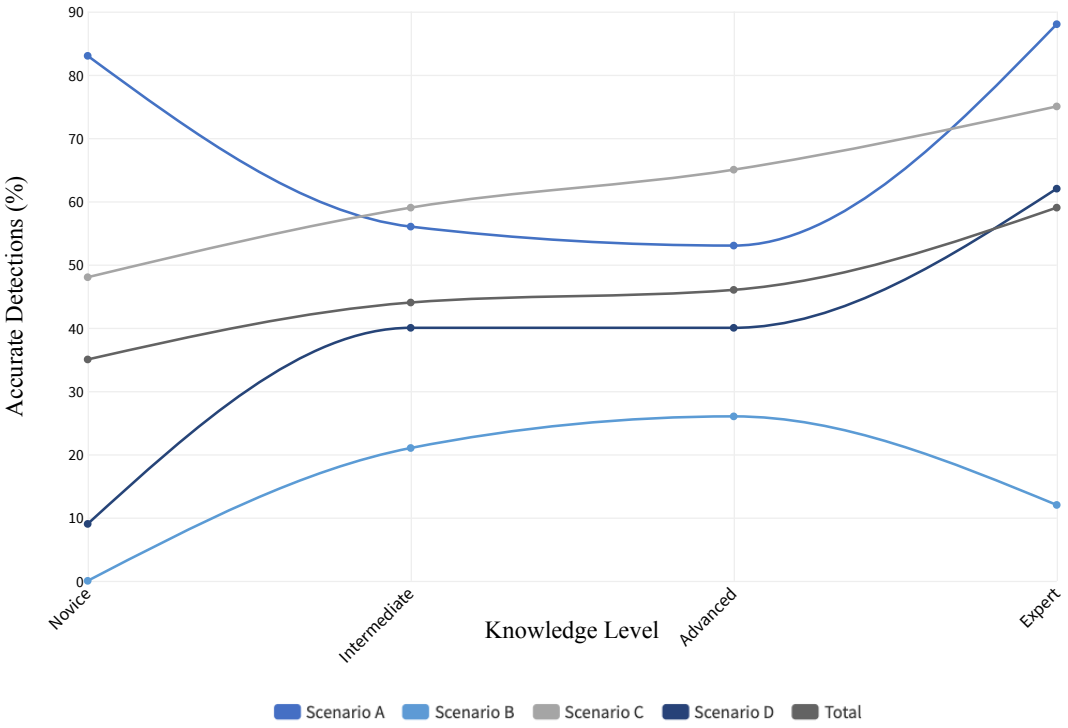


Figure 5.7: Accuracy within Knowledge levels

However, when the four scenarios were considered separately a clear dichotomy was found between scenario *A*-*B* and *C*-*D*. Results in Table 5.6 show that for scenario *A*, where no attack was performed, detection accuracy across the four knowledge levels did not demonstrate any association between knowledge level and ability to detect an infected device. Novice (83%) and Expert (88%) demonstrated similar accuracy, and better than that of both Intermediate (56%) and Advanced (53%). In Table 5.7, the results for scenario *B* again show that detection accuracy across the four knowledge levels did not demonstrate any association between knowledge level and ability to detect an infected device. Participants across all knowledge levels reported finding it difficult to identify an infected device from just the live video feed.

> *"I couldn't see anything happen so assume they were not infected", "it wasn't possible to tell if anything bad happened"* (Novice Participant)

> *"I do not think these cameras were infected, as I expected some stuttering or a black out, but this did not happen"* (Intermediate Participant)

> *"There was no stuttering or black outs of video, so I would say neither camera was infected"* (Advanced Participant)

> *"I could not tell at all if the cameras were infected. I only noticed a timing*

difference between the two videos, concerning the L/R letters animation" (Expert Participant)

For scenarios *C-D* a clear association was found between the two variables, as shown in Figure 5.7. As technical knowledge increased from Novice to Expert, so did participants ability to detect an infected device: Novice (48%), Intermediate (59%), Advanced (65%) and Expert (75%) (see Table 5.8). Presenting network communication as shown in Figure 5.5a appeared to greatly improve their awareness of a threat, and ability to correctly detect when an attack took place. Participants reported colour being helpful and a positive contributing factor to better detection accuracy:

> *"Program code went red"*, *"Bulk black lines appeared"*, *"Maybe the black bits with red writing may be something bad?"* (Novice Participants)

> *"yes wire shark made it easier to see that it was infected by all the random traffic"*, *"there were red warnings on the screen"*, *"Vast number of red highlighted addresses"* (Intermediate Participants)

> *"On the first the red warning messages were visible"*, *"I saw a lot of areas highlighted in red, red highlights usually denotes a problem, so by deduction, those were errors"*, *"Red text black blocks"* (Expert Participants)

It was clear from participants comments that the way information is presented, and importantly the colours used, helped to aid better detection. This was evident even among Novice participants, who appeared not to fully understand what the information was showing, but were able to use it to become more situationally aware of what was happening with the IoT device.

> *"Red normally represents danger, so I would guess the parts of video which were red was when the cameras were infected"*, *"It was difficult to tell but I would guess the bits which flashed red (first camera) and the bits that flashed black/red (both cameras) could be a warning of something happening ?"* (Novice Participants)

In scenario D, participants were again shown network communication as shown in Figure 5.5b. Results in Table 5.9 again demonstrated an association between the two variables, although not as strong as the previous scenario. Data presentation differed from the network traffic in scenario *C*, and appeared to be a contributing factor in detection rates, particularly within the Novice knowledge group where the detection rate significantly dropped to (9%).

From the analysis in this study it is possible to conclude that the authors assertion in [3] that "a lack of technical knowledge, and the ability to explore network communication, results in little or no awareness of security issues", is true in part. The results in this study show that a lack of network communication can result in little or no awareness of security issues; however, if presented with data, awareness can be improved. Presentation of the data is however also vitally important, otherwise the presence of the additional data, can have little impact. Since the results were not statistically significant in all four scenarios, the null hypothesis $\mathbf{H_1}$ cannot be fully rejected.

This study also tested the assumption that age has an impact on security [190] to discover if there was a relationship between the studies dependent variable (accuracy) and participant's age which was used as the second independent variable in this study. To test this assumption the hypothesis was defined in section 5.2.1 as:

> **$\mathbf{H_2}$:** There is no association between **detection accuracy** and **age** when detecting threats in consumer networks.

Accuracy levels for each age group were summarised in Table 5.15 and did not indicate an association existed between age and ability to detect an attack. Results were also not found to be significant for any scenarios, therefore the null hypothesis $\mathbf{H_2}$ was accepted and the study concludes there is no association between age and ability to detect threats in consumer IoT networks. This chapter set out to answer the question "*Can users visually detect the presence of threats within consumer IoT networks ?*". It is clear from the results that users find it very difficult to detect the presence of threats, however, this can be improved through the presentation of additional information, which will be explored in the next chapter.

5.5 Conclusions

This chapter examined the awareness and perception of threats found within the IoT. First, it explored how users value and perceive security and privacy in smart devices. It also analysed user requirements from IoT devices, and assessed their ability to detect threats, in the context of demographic characteristics (technical knowledge and age). It is still unclear whether a clear association exists between demographic characteristics and the ability to detect threats. However, the results suggested that users valued security and privacy but found identifying threats difficult. The study found that a lack of network communication can result in little or no awareness of security issues; however, if presented with data, awareness can be improved. Presentation of the data is, however, vitally important, otherwise the presence of the additional data, can have little impact. This evidence provides the justification for the further research presented in subsequent chapters, namely an exploration of different modalities for presenting data to non technical users. The next chapter reports the results of a study which examined the use of conversational agents for improving situational awareness. The development of two agents based on **Aural**$^{(\mathrm{Au})}$ and **Verbal**$^{(\mathrm{Ve})}$ modalities is presented, and the findings of a cross-sectional viability study reported.

Chapter 6

Cross-sectional Study to Test the Viability of Conversational Agents to Improve Cyber Situational Awareness

Chapter 5 examined how users value and perceive security and privacy in smart devices found within the IoT. The study analysed user requirements from IoT devices, and the importance placed upon security and privacy. It also assessed a users ability to detect threats, in the context of technical knowledge and experience. The study showed that although users reported to value security and privacy they could not adequately detect when an IoT device was infected and performing attacks. This chapter seeks to address this issue and examine the use of conversational agents for improving situational awareness. The chapter presents the development of two agents based on **Aural**$^{(Au)}$ and **Verbal**$^{(Ve)}$ modalities and reports the findings of a cross-sectional viability study.

6.1 Introduction

The focus of this chapter is attempting to answer sub research question SQ3: "*Are conversational agents a viable method for making users aware of threats in consumer IoT networks?*". Firstly, a quantitative approach was used to examine the use of two conversational agents for improving Situational Awareness of threats in consumer IoT networks. Mica Endsley's [1] model was adopted to assess how participants perceive device activity, comprehend this in the context of their environment, and use the knowledge to determine if a threat exists. Secondly, a qualitative approach was

taken to examine feedback about the conversation agents using a thematic analysis technique. This twofold approach to analysis was carried out in order to examine the viability of the conversational agents for improving situational awareness, and to elicit feedback which could inform possible refinements to the conversational agents ahead of a longitudinal study in Chapter 7.

6.2 Methodology

6.2.1 Experimental Variables

This study measured participants ability to be situationally aware of threats within consumer IoT networks using two conversational agents. The statements created in Section 3.3.5, which had previously been mapped to Endsley's SA model [1], were used to calculate a Cyber Situational Awareness score (*CSAS*). Since the aim of this study was to test the overall viability of the agents, the three layers of Endsley's model were combined to create a single metric (*CSAS*). This metric was used as the dependent variable and was derived from calculating the sum of the three means **Perception**, **Comprehension**, and **Projection** using:

$$\text{CSAS} = \frac{1}{3}\sum_{i=1}^{3}\mu\ (\text{X}_\text{i})$$

Where:
X_1: $\sum(pe_1, pe_2, pe_3)$
X_2: $\sum(co_1, co_2, co_3)$
X_3: $\sum(pr_1, pr_2, pr_3)$

Data was collected using Likert scales and was therefore ordinal. The **Aural**$^{(Au)}$ and **Verbal**$^{(Ve)}$ conversational agents were used as the independent variables. Since the aim of the study to test the assumption that participants awareness of threats could be improved using the agents, a Pre-Study/Post-Study analysis of the dependent variable was performed to measure any reported differences in (*CSAS*). The hypothesis for this study is therefore defined as:

Hypothesis A: Situational Awareness of threats will be improved using conversational agents.

The previous study, presented in chapter 5, found that a lack of understanding/knowledge of network communication can result in little or no awareness of security issues; however, if users are provided with additional information in the correct format, situational awareness of threats can be improved. A visual modality was used to present

the additional information to non technical users and was demonstrated to be a convenient and effective method of improving awareness. In addition, previous research [143, 140, 141] has also suggested awareness and understanding can be improved using other learning modalities. Therefore, the aim of this study was to explore if conversational agents, based on additional learning modalities, could improve situational awareness of threats. Specifically, the study tested the assumption that conversational agents based on **Aural**$^{(Au)}$ and **Verbal**$^{(Ve)}$ modalities could be used to present additional information to users to improve situational awareness of threats. It was hypothesised:

> **H_1:** Post-Study **Cyber Situational Awareness Score** (*CSAS*) of the ***Intervention*** group will be higher than the Pre-Study **Cyber Situational Awareness Score**.
>
> **H_2:** Post-Study **Cyber Situational Awareness Score** (*CSAS*) of the ***Control*** group will be lower than the Pre-Study **Cyber Situational Awareness Score**.
>
> **H_3:** Post-Study **Cyber Situational Awareness Score** (*CSAS*) of the ***Control*** group will **Not** be significantly different than the ***Intervention*** group Pre-Study **Cyber Situational Awareness Score**.

The first hypothesis (**H_1**) was used to measure if participants became more situationally aware of threats as a result of using the conversational agents. The results in Chapter 5 and research in [19] suggested that participants may have some awareness of potential threats, but may not fully understand the implications or potential for misuse. The second hypothesis (**H_2**) was, therefore, used to measure changes in situational awareness, as a result of watching a video which explained how insecure smart devices could be used to perform attacks on the internet. If the control groups Post-Study results were lower than their Pre-Study results, this may confirm the existence of this phenomenon. Finally, it was important to establish that any measured changes in situational awareness could be attributed to the use of the agents and not as a result of any hidden factors [149], such as the video which was watched. Participants were therefore split into two groups and the Post-Study ***Control*** group results compared with the Pre-Study *Intervention* group results in hypothesis (**H_3**). If there was no significant difference found between the Post-Study *Control* group results and the Pre-Study *Intervention* group, the reliability of the results could be established and the mitigation of confounding variables confirmed.

6.2.2 Study Design

A cross-sectional study was conducted to evaluate the viability of using conversational agents to improve Cyber Situational Awareness (See Figure 6.1). Participants were recruited and assigned to two groups as described in Section 6.2.4. Participants in the *Intervention* group watched an introductory video[1] about threats facing smart homes and completed a short related activity. The video was chosen since it provided a good summary of recent DDoS attacks which have targeted smart devices found within the IoT, and could, therefore, be used to gauge participants current awareness of such threats. In addition, it could be used to check for confounding variables as discussed below. The participants were then asked to complete a Pre-study survey and indicate their level of agreement with nine statements (See Table 6.1) relating to their awareness and ability to monitor smart device and network activity. These were mapped to Mica Endsley's SA model as described in Section 3.3.5. Participants were then asked to use the agents to answer questions relating to use-case scenarios presented in Section 6.2.2, and provide feedback on their effectiveness. Finally, the participants completed a Post-study survey of the same nine statements, and variance in their attitudes were recorded. The use of a *control* group enabled confounding variables to be mitigated and enhanced scientific rigour. Participants in the *control* group completed the same Pre-Study survey, before watching the introductory video about threats facing smart homes and completing the short related activity. Finally, they completed the Post-Study survey. This approach allowed their Pre-Study/Post-Study scores for the nine confidence statements to be compared and variance in the scores measured. If they reported being less confident in the Post-Study survey, this may indicate they initially reported inaccurate confidence levels due to a lack of awareness of the threats facing smart home environments. The study design also allowed the *intervention* group results to be checked for confounding variables, by comparing the *control* group Post-Study scores with the *intervention* group Pre-Study scores. Since these were both completed after watching the short video and activity, they should be similar and validate the results of the study. Pre-Study/Post-Study survey are presented in Appendix I.

[1] https://tinyurl.com/ResourcesCMcD

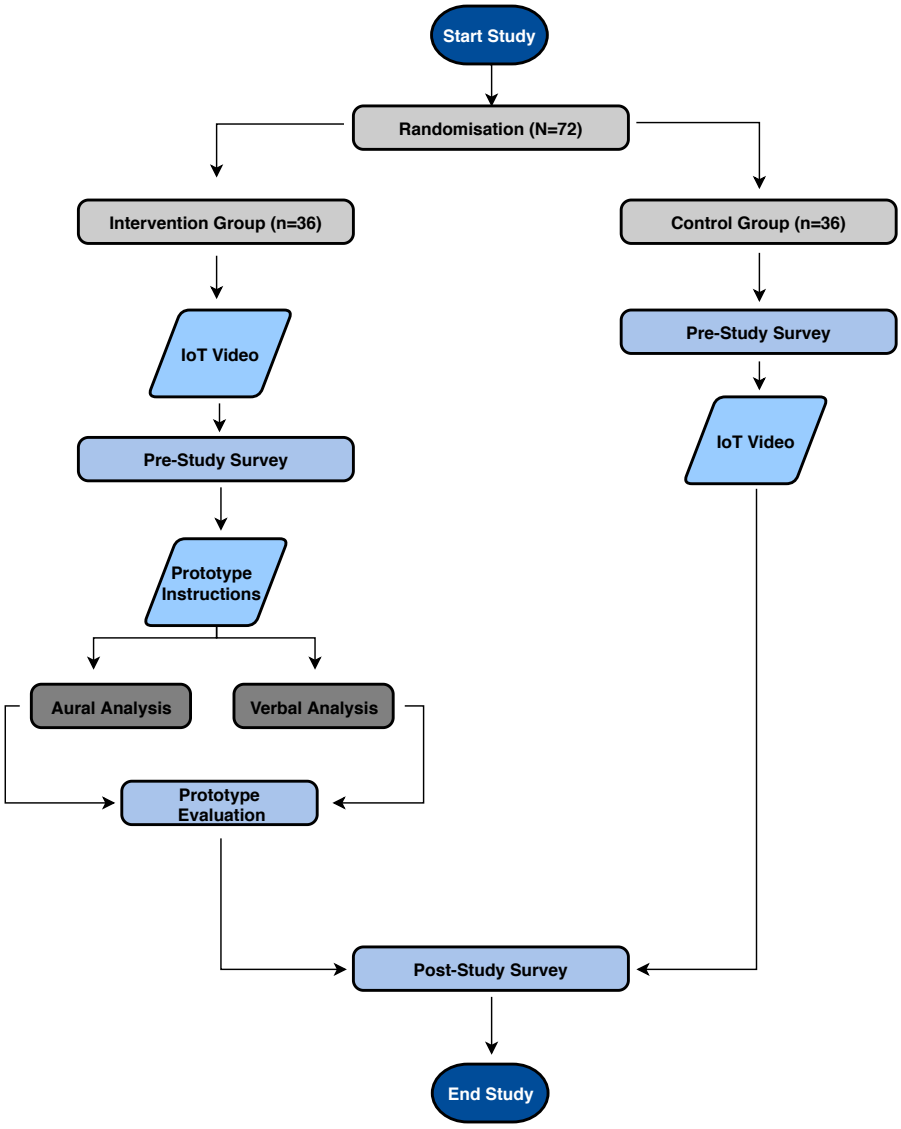


Figure 6.1: Cross Sectional Study Design.

Use-Case Scenario Design

To evaluate the viability of using conversational agents to improve Cyber Situational Awareness the use-cases developed in Section 3.3.2 were used. Each use-case was designed to represent a realistic example of how a user might use the agents to monitor smart device and network activity. To give use-cases context and ensure participants had a realistic experience of monitoring smart device and network activity, a contextual scenario was created for each use-case. Table 3.1 describes the five use-cases used in the Pilot and the further two use-cases used in the Main study. The corresponding scenario devised for each use-case is presented in Appendix D.1. Finally, Table 3.3 demonstrates how each scenario mapped to the corresponding use-case in the study.

Cyber Situational Awareness

Cyber Situational Awareness was measured using the nine confidence statements created in Section 3.3.5. A survey was generated which included the nine confidence statements, plus seven additional statements relating to general smart device security (See Table 6.1). The additional general security statements were only used as a distraction [200], to disguise the focus of the nine confidence statements. This was important as previous studies have emphasised the importance of avoiding the introduction of bias into a study by priming participants [201, 202]. The use of the seven additional statements ensured participants were unaware that the focus of the study was on their confidence to detect threats, which could have resulted in inflated confidence levels being reported. For each statement in the survey participants were asked to indicate their level of agreement using a five-point likert scale from *Strongly Disagree (1)* to *Strongly Agree (5)*. Finally, participants completed the same survey before and after using the conversational agents, and the differences in their responses was compared. The Pre-Study/Post-Study Survey is presented in Appendix I.

Table 6.1: Pre-study/Post-study Cyber Situational Awareness (CSA) Statements

Five-point Likert Scale from Strongly Disagree (1) to Strongly Agree (5)

#	**SA Statement**
pe1	I am confident I can tell which smart devices are using my home network.
±	Smart devices are more secure than non smart equivalent devices.
pe2	I am confident I can tell how often a smart device is communicating on my homework, and how much of the available network bandwidth it is using.
±	Smart devices update themselves automatically.
pe3	I am confident I can tell which smart devices have the highest usage on my home network.
±	Smart devices are intelligent and can protect themselves from attackers.
co1	I am confident I can tell if my network is experiencing a normal level of device communications and bandwidth usage.
±	Smart devices alert you if an attacker is trying to compromise the device.
co2	I am confident I can tell if a smart device is functioning normally.
±	Smart devices are less likely to be targeted by attackers.
co3	I am confident I can tell if a smart device is using my home network more or less than normal.
pr1	I am confident I can tell if an attack has taken place on my home network.
±	Smart devices in the home are not accessible from the Internet.
pr2	I am confident I can tell if a smart device on my home network has been compromised.
±	Smart devices in the home can be used to perform attacks on the internet.
pr3	I am confident I could tell in the future if my home network or smart device had been compromised.

± *used as a distractor statement*

6.2.3 Conversational Agent Design

This section describes the design of the **Aural**$^{(Au)}$ agent used in both the initial Pilot study (See Section 6.2.5) and the main study, and the additional **Verbal**$^{(Ve)}$ agent used in the main study only. A more detailed description of each agent development is presented in Appendix B.

ETL Pipeline

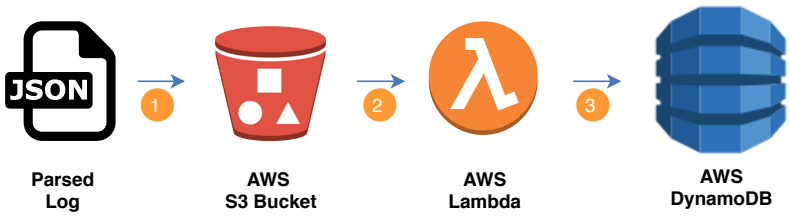


Figure 6.2: IDS Log Parsing and Storage.

In Chapter 4, 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 to classify traffic as either normal or attacks associated with the *Mirai* malware. A subset of the dataset was used in this study, and contained 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 Chapter 4, the complexity of the information meant they had limited value when using the conversational agents. Since the features would not be required later, they were removed. Finally, the *csv* files were concatenated, converted to *JSON* format, and stored in a specified directory ready to be ingested by the ETL pipeline. A sample record from the newly amended dataset is found in Source Code 6.1.

```
1 {
2  "ID": "487",
3  "DateTime": "20/01/2019 19:01",
4  "SourceDevice": "192.168.252.40",
5  "DestinationDevice": "180.130.236.179",
6  "DataType": "TCP",
7  "Activity": "normal"
8 }
```

Source Code 6.1: Sample JSON record

The ETL pipeline to *Extract*, *Transform*, and *Load* IDS logs into *DynamoDB* is presented in Figure 6.2. Suitable IDS logs are parsed and stored in the specified directory, ready to be ingested by the ETL pipeline. For this study, the IDS logs consisted of the classified dataset (subset) in *JSON* format. 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, ready to be queried by an agent.

Aural Agent

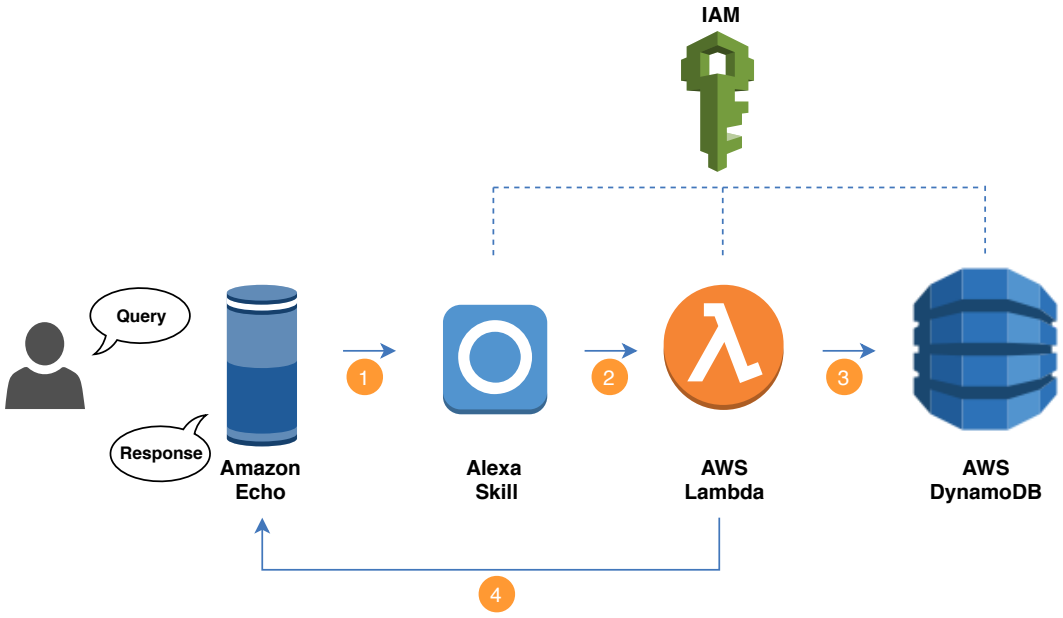


Figure 6.3: Aural Agent Architecture.

Primary input for the **Aural**$^{(Au)}$ conversational agent is speech derived from Amazon Alexa enabled devices. Input is analysed using natural language processing (NLP) techniques to understand the user query (*intent*), and is matched to sample phrases (*utterances*) of ways a user could ask the query. The intent is used to query the secondary input source (IDS logs stored in a *DynamoDB* table) for an appropriate answer to the query, and responses are returned accordingly.

The agent consists of three main components: a database of classified IoT traffic created in Chapter 4, NLP engine as an interface between a user and the Alexa device, and a query handler. The **Aural**$^{(Au)}$ agent architecture is presented in Figure 6.3. The speech recognition engine is contained in the Alexa device, the query handler is the developed *Alexa Skill* and *AWS Lambda function*. Finally, a *DynamoDB* database is used to store and query classified IDS logs.

In Figure 6.3, 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)*, matches them to specified (*utterances*) and configured functionality *intents* and forwards the request to the *AWS Lambda function (step 2)*. Next, a query request is triggered to interact with the *DynamoDB* table *(step 3)*. Once fulfilled, an appropriate answer to the user query is returned. Finally, the Alexa skill generates an aural response from the returned answer, invokes the Echo device, which in turn communicates the response 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 also hosted on AWS Lambda, using server-less technology to allow 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*.

The *intents* developed and used by the **Aural**$^{(Au)}$ agent are described in Section 6.2.3 and were tested using the *use-cases* described in Section 6.2.2. An example user query and agent response is shown in Figure 6.5(a).

Verbal Agent

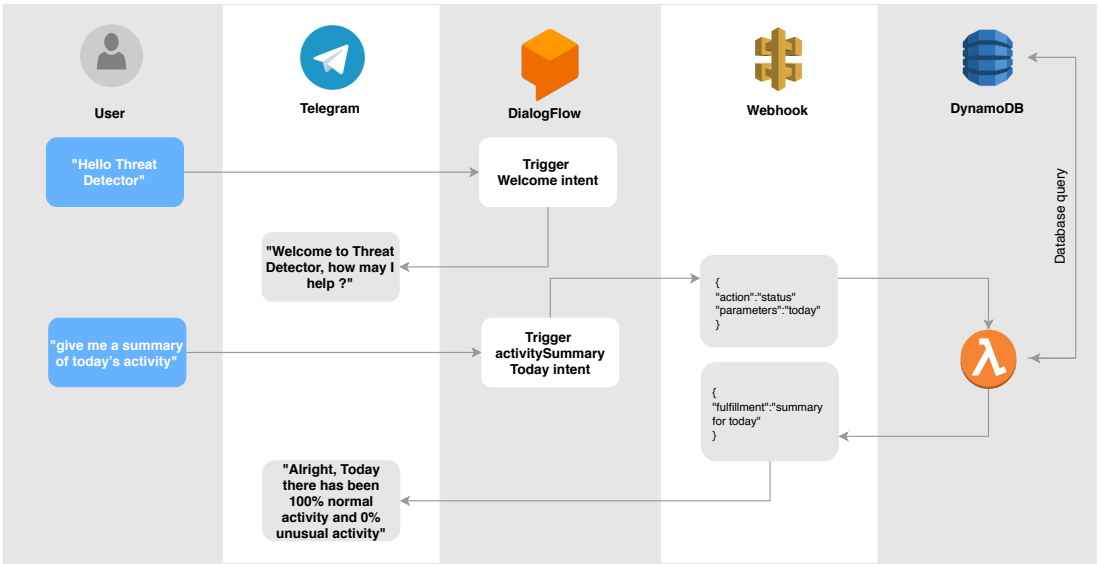


Figure 6.4: Verbal Agent Architecture

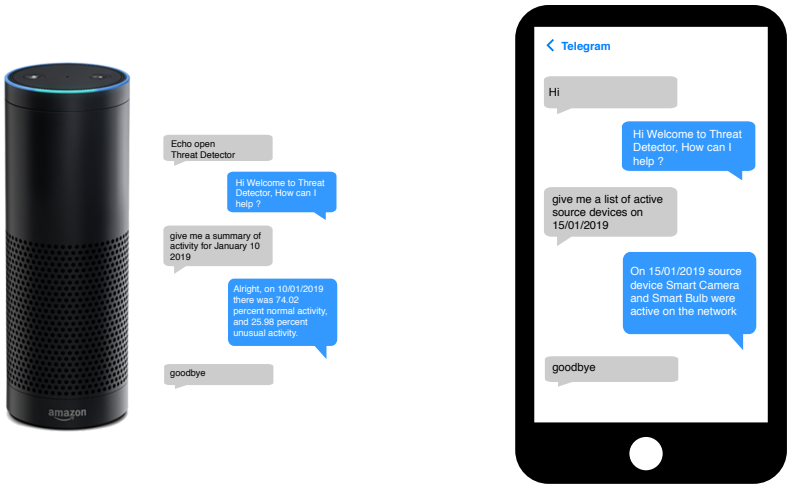
Primary input for the **Verbal**$^{(Ve)}$ conversational agent is text derived from a chatbot running on the *Telegram*[1] messaging service. The architecture is similar to that of the **Aural**$^{(Au)}$ agent, however, is hosted on Google's *Dialogflow*[2], a popular platform for creating conversational agents. Natural language processing (NLP) techniques are also used to understand user queries and extract information from a *DynamoDB* table to identify threats which may have occured.

The process of interaction between a user and the **Verbal**$^{(Ve)}$ agent is shown in Figure 6.4. A user interacts with the agent using a chatbot deployed in the *Telegram* app, running on a mobile phone or tablet. The user query (*intent*) is sent from *Telegram* to *Dialogflow* which uses Google's Machine Learning NLP to understand the user query. For each intent, sample phrases (*utterances*) of what a user might say when interacting with the agent were provided. Dialogflow then uses extensive accumulated domain knowledge to analyse and understand the user's intent, to ensure accurate query responses. Once understood the query is forwarded to the *AWS Lambda Function* however, since the query requires to be forwarded from Google's platform to Amazon, an *AWS API* was created and used as a webhook, creating a bridge between the two platforms. Whenever a user query is received by the *Lambda function*, the query handler is triggered and interacts with *DynamoDB* to find a suitable response the user query. The response is sent back to Dialogflow and is then forwarded to the chatbot running in *Telegram*.

The **Verbal**$^{(Ve)}$ agent uses the same backend infrastructure as the **Aural**$^{(Au)}$ agent which is hosted on AWS infrastructure as a scalable serverless solution.

The *intents* developed and used by the **Verbal**$^{(Ve)}$ agent are described in Section 6.2.3 and were tested using the *use-cases* described in Section 6.2.2. An example user query and agent response is shown in Figure 6.5(b).

[1] https://telegram.org/
[2] https://dialogflow.cloud.google.com/



(**a**) Aural Agent (**b**) Verbal Agent

Figure 6.5: Example Agent Conversations

Agent Intents

Conversational agents built with Amazon's *Alexa Skills Kit*[1] or Google's *Dialogflow*[2] use *intents* to represent an action that fulfills a user's request (Aural or Verbal). Ten custom intents were configured, and used to trigger specific event functionality and enable a user to query the IDS logs in *DynamoDB* for information. Seven in-built Amazon intents were also used as triggers to perform preconfigured functionality such as *repeat*, *stop* or *cancel* an intent.

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. For example, 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 "give me summary details 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 represented variable information specified by a user, were assigned a *slot/action*. 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*

[1]https://developer.amazon.com/en-GB/alexa
[2]https://dialogflow.cloud.google.com/

six. In total ten intents were created. Five were used in the initial pilot study, and a further five were added in the main study 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.
6. ***networkStatusToday:*** Responds to a user query and informs if there has been any issues detected on the network.
7. ***listSrcDevToday:*** Responds to a user query and returns a list of all active source devices on the network today.
8. ***listSrcDevByDate:*** Responds to a user query and returns a list of all active source devices on a specified date.
9. ***activityTotalBySrcDevLastThreeDays:*** Responds to a user query and returns details of how much activity a specified source device had on each of the last three days.
10. ***activityTotalLastThreeDays:*** Responds to a user query and returns a summary of how much activity has occurred on each of the last three days.

Example intents, utterances and responses are presented in Table 6.2, and full details of each intent is provided in Appendix C.

Table 6.2: Example Agent Intents, Utterances and Responses

	Intent	Utterance	Response
i1	activitySummaryToday	(1) "show a summary of today's activity" (2) "give me a summary of today's activity"	Alright, Today there has been 66.19 percent normal activity and 33.81 percent unusual activity. Anything else I may help you with?
i4	firstUnusualActivityByDate	(1) "give me detail of first {input} activity on {unusualDate}" (2) "what was the first {input} activity on {unusualDate}" (3) "on {unusualDate} what was the first {input} activity"	Alright, First unusual activity on 2019-01-20 For which I.D is 690, Date Time is 20/01/2019 09:53, Source Device is Smart Camera, Destination Device is Smart Fridge, Data Type is UDP, Activity type is unusual. Anything else I may help you with?
i5	activityDetailsByID	(1) "give me details of activity id {ID}" (2) "show me activity id {ID}" (3) "what is the details of activity id {ID}"	I.D is 366, Date for that is 2019-01-18, time is 19:39, Source Device is Amazon Echo, Destination Device is Smart Camera, Data Type is ARP, Activity type is normal. What else would you like to know?

Full list of intents are presented in Appendix C

6.2.4 Participants

In the main study the aim was to assess the viability of conversational agents for improving awareness of threats facing smart devices. Convenience sampling was employed, with participants selected due to their convenient accessibility, and proximity to the author. Informed consent was provided by participants by reading the study agreement on the first page, before indicating their consent to participate by proceeding to the next page of the survey. Participants were recruited at university Applicant Day events in February and March 2019, and during public events held in the university during British Science Week 2019. This approach enabled us to collect a wide range of views and avoid oversampling of a specific demographic, namely the local student population. A total of eighty participants started the study, and were randomly assigned to either the *control* or *intervention* group creating groups of similar sizes. Eight participants did not complete the study resulting in an attrition rate of 10%. When attrition occurs, the groups can become dissimilar which can lead to bias in the estimated effect

of the intervention [203]. However, although this study had high differential attrition (13.5%), the attrition predominately took place in the control group, therefore the intervention results are still considered to have validity. Although the intention was to avoid oversampling, the largest demographic was participants aged 18-24 with advanced technical skills as presented in Table 6.3. This however was considered acceptable since statistically this is the largest demographic of conversational agent users [204].

Table 6.3: Participant Demographic

Age		%	**Ability**		%
18-24	34	(47)	**Novice**	6	(8)
25-39	25	(35)	**Intermediate**	20	(28)
40-59	11	(15)	**Advanced**	32	(44)
60+	2	(3)	**Expert**	14	(20)

n = 72

Randomisation is an important element of a well-designed experiment [157]. Participants were therefore randomly assigned to the two groups (Control and Intervention) as shown in Table 6.4.

Table 6.4: Group Allocation

	Novice	**Intermediate**	**Advanced**	**Expert**
Control	**3** *p23, p55, p58*	**5** *p43, p46, p59, p64, p66*	**19** *p4, p5, p9, p10, p11, p13, p16, p24, p25, p26, p37, p42, p67, p70, p71, p74, p76*	**9** *p1, p2, p22, p29, p31, p38, p48, p54, p68, p69, p80*
Intervention	**3** *p56, p57, p60*	**15** *p14, p21, p39, p40, p41, p44, p45, p49, p50, p51, p53, p62, p63, p72, p73*	**13** *6, p7, p8, p12, p15, p18, p19, p33, p47, p65, p75, p77, p79*	**5** *p27, p28, p30, p34, p36*

n = 72

6.2.5 Pilot

To test the effectiveness of the prototype conversational agent a pilot study was conducted in December 2018, using a representative sample (n=12) of users. This size adhered to recommended guidelines [205] [206] of being a minimum of 10 particpants or 10% of the treatment group in the main study. 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 evaluate the conversational agent, each participant was asked to complete a Pre-Study survey and indicate their level of agreement with five statements (See Table 3.4) relating to their awareness and ability to monitor smart device and network activity. A Likert-type scale from *Strongly Disagree* to *Strongly Agree* was used. To test the functionality of the conversational agent, participants were then asked to use the agent to answer questions relating to *five* use-cases presented in Section 3.3.2. The use-cases were designed to represent realistic descriptions of how a user might want to use the conversational agent for monitoring smart device and network activity. Finally, participants completed a Post-Study survey of the same five statements, and variance in their attitudes was recorded. Since related groups were being compared, where participants completed the same survey Pre and Post study, Wilcoxon signed-rank tests were used for comparisons. A statistically significant median increase ($p < .05$) was observed in three statements [20]. This clearly demonstrated the agent had the potential to make a positive contribution towards improving situational awareness of threats in smart homes. On conclusion of the Post-Study survey, participants were asked for suggestions of how the agent could be improved. These can be summarised as follows:

1. The ability to get a simple status of the network and if any unusual activity has occurred.
2. The ability to see which devices have been active on the network on a given date, and their total activity.
3. The ability to see the total activity for a device, and combined total for the network.

Amendments for Main Study.

The feedback was used to make amendments and additions to the **Aural**$^{(\mathrm{Au})}$ agent to improve the utility of the agent for the main study. Firstly, feedback suggested the functionality of the agent should be extended, therefore *five* additional intents were added to the agent for the main study (See Appendix C.1). This also resulted in the addition of *two* extra use-cases (See Table 3.1). Finally, participants found it

challenging to find devices using their IP address and suggested using the device's name "don't use ip address to search for a device but rather its name" (*# p12*). Therefore, all intents and agent responses were amended to use device names.

6.3 Results

6.3.1 Cyber Situational Awareness Score (CSAS)

Participants awareness and confidence to detect threats was measured and is presented in Table 6.5. Pre-Study/Post-Study Mean, Median and standard deviation for both the *control* and *intervention* groups are reported. In addition, the three layers of Endsley's SA model were combined and are reported as a single metric (*CSAS*). This was derived from the sum of the three means **Perception**, **Comprehension**, and **Projection**.

Table 6.5: Mean, Median and Standard Deviation: Cyber Situational Awareness Score (CSAS)

		SA Levels					**CSAS**
		$\overline{x}$	$\tilde{x}$	σ	min	max	$\sum$
CTL	**Pre-Study**	3.50	3.33	.69	2.11	5.00	10.51
	Post-Study	2.46	2.50	.39	1.56	3.44	7.37
INT	**Pre-Study**	2.41	2.33	.41	1.56	3.44	7.23
	Post-Study	4.06	4.00	.44	2.78	4.78	12.17

To check if differences between Pre and Post-Study scores were significant Wilcoxon Signed-Rank tests were carried out. The Post-Study *Intervention* median score (4.00) was found to be higher than the Pre-Study score (2.33). A Wilcoxon Signed-Rank test indicated that the difference was statistically significant, Z= -5.240, $p < .001$. The Post-Study *Control* median score (2.50) was found to be lower than the Pre-Study score (3.33). A Wilcoxon Signed-Rank test indicated that the difference was statistically significant, Z= -5.076, $p < .001$.

To test the validity of the data a Cronbach alpha coefficient test was performed on both groups Pre and Post-Study situational awareness scores. Cronbach alpha coefficient for the *Control* and *Intervention* groups ($\alpha = .77, .84$ respectively) were found to be above the recommended value .70 [207], therefore the data was considered valid with good internal consistency.

6.3.2 Conversational Agent Effectiveness

Participants effectiveness rating for both conversational agents was measured and is presented in Table 6.6. The Mean and Median effectiveness scores were calculated for each of the seven use-case scenarios and used to compare the two agents. Results show that the effectiveness of the **Aural**$^{(Au)}$ agent was rated higher in three scenarios (*sc1 (4.56), sc2 (4.36), sc3 (4.47)*), while the **Verbal**$^{(Verbal)}$ agent was rated higher in the remaining four scenarios (*sc4 (3.94), sc5 (3.97), sc6 (4.31), sc7 (4.22)*). However, in the three scenarios where the **Aural**$^{(Au)}$ agent was rated higher, the mean scores for the two agents were comparable: *sc1* (.23+), *sc2* (.08+), *sc3* (.14+). In comparison, the mean difference between agents in the scenarios where the **Verbal**$^{(Verbal)}$ agent was rated highest was shown to be more significant: *sc4* (1.05+), *sc5* (1.11+), *sc6* (.62+), *sc7* (.55+).

Table 6.6: Mean, Median: Agent Effectiveness

		$\overline{x}$	$\tilde{x}$	min	max
sc1	**Au**	4.56	5.00	3.00	5.00
	Ve	4.33	4.00	3.00	5.00
sc2	**Au**	4.36	4.00	3.00	5.00
	Ve	4.28	4.00	3.00	5.00
sc3	**Au**	4.47	5.00	3.00	5.00
	Ve	4.33	4.00	3.00	5.00
sc4	**Au**	2.89	3.00	1.00	5.00
	Ve	3.94	4.00	2.00	5.00
sc5	**Au**	2.86	3.00	1.00	5.00
	Ve	3.97	4.00	3.00	5.00
sc6	**Au**	3.69	4.00	2.00	5.00
	Ve	4.31	4.00	3.00	5.00
sc7	**Au**	3.67	4.00	2.00	5.00
	Ve	4.22	4.00	3.00	5.00

Aural$^{(Au)}$ ***Verbal***$^{(Ve)}$

$n = 36$

Figure 6.6 further demonstrates the difference in effectiveness rating between the two agents. Participants rated the agents on a scale of One to Five (1 being low and 5 high). Results show that for the first three scenarios the reported effectiveness scores

are comparable, with 88.89% of participants rating both agents either 4/5 or 5/5. For scenarios four and five the difference in effectiveness rating is more evident with 69.45%, 72.22% respectively rating the **Verbal**$^{(Ve)}$ agent either 4/5 or 5/5. In comparison, for scenarios four and five participants rating the **Aural**$^{(Au)}$ agent either 4 or 5/5 was only 27.77%, 27.78%. Finally, for scenarios six and seven the **Verbal**$^{(Ve)}$ agent was clearly rated much higher with 94.44%, 88.89% rating the agent as either 4/5 or 5/5. In comparison, the **Aural**$^{(Au)}$ agent was only rated as either 4/5 or 5/5 by 58.34%, 55.56% of participants respectively.

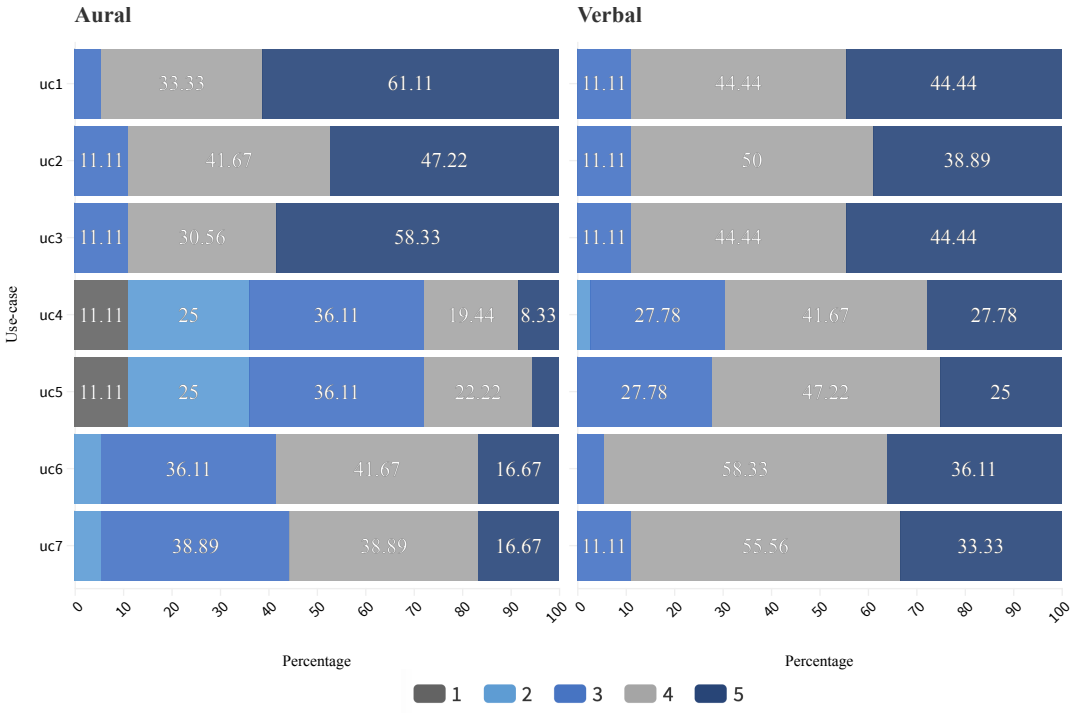


Figure 6.6: Aural and Verbal Effectiveness Rating

6.4 Discussion

This study was undertaken to assess the viability of using conversational agents to improve Cyber Situational Awareness. Specifically, it sought to explore if user awareness and perception of threats facing consumer IoT could be improved through using the agents which were developed. As previously mentioned, participants in the *Intervention* group were asked to use the two conversational agents to answer questions relating to use-case scenarios presented in Section 6.2.2, and provide feedback on their use and effectiveness. Feedback was later used to refine the agents for use in the longitudinal study in Chapter 7. Template analysis was used to analyse the feedback data as

described in Section 3.4.2. Coding tables produced from the qualitative analysis of responses are presented in Appendix E. For brevity, a representative sample of responses are provided below for each of the questions asked. Participants were first asked what they liked and disliked about each of the conversational agents. Participants reported the usability of both agents to be a strength:

> *"Convenient to quickly check if devices are ok"* (# *p1*)
>
> *"handy way to check your devices are ok"* (# *p17*)
>
> *"much quicker than checking each device individually"* (# *p9*)
>
> *"that I can check my devices are ok from my phone anytime I want"* (# *p3*)

Participants reported finding the agents a convenient and easy way to quickly check if their devices were functioning normally. The portability of the **Verbal**$^{(Ve)}$ agent was found to be particularly useful since it allowed devices to be checked outside of the home. Participants also reported finding the interactive nature of the agents to be enjoyable and educational.

> *"I enjoyed using this technology"* (# *p7*)
>
> *"It was actually quite fun to use the Alexa for checking smart devices"* (# *p22*)
>
> *"I liked learning new technology"* (# *p29*)

Interestingly, elements of agent usability were also reported to be areas least liked about the agents. Participants reported difficulty with the **Aural**$^{(Au)}$ agent understanding their voice commands. While typing difficulties were reported with the **Verbal**$^{(Ve)}$ agent.

> *"Was a bit laggy for a while"* (# *p3*)
>
> *"sometimes struggled to understand me first time"* (# *p13*)
>
> *"had to speak like a robot for it to understand me"* (# *p17*)
>
> *"typing the same questions multiple times may become tedious"* (# *p10*)
>
> *"typing each question could be prone to error"* (# *p16*)
>
> *"lack of help to know what queries were available"* (# *p34*)

Since the intention of soliciting feedback was to inform refinements needed in preparation for the final longitudinal study in Chapter 7, participants were next asked to

recommend one improvement that would make the agents better. As speech recognition was identified as an issue with the **Aural**$^{(Au)}$ agent, this was an area recommended for improvement.

> *"improve how it recognises peoples voice"* (*# p13*)

> *"understand Scottish accents better"* (*# p32*)

In addition, since participants reported issues relating to typing when using the **Verbal**$^{(Ve)}$ agent it was recommended to improve this area by enabling predictive typing and providing a list of available queries which can be used.

> *"help facility to see which questions are available and what they do"* (*# p17*)

> *"some kind of reminder of the the questions that are available"* (*# p34*)

> *"predictive typing like when you are texting on your phone"* (*# p23*)

Finally, it was recommended that more queries be added to both agents, in particularly a method to provide a quick summary of activity for multiple days. It was recommended that the availability of the **Verbal**$^{(Ve)}$ agent should be expanded to other popular platforms such as Whatsapp and Facebook Messenger.

> *"quick summary for multiple days"* (*# p10*)

> *"never heard of Telegram, more people might use it if it was on something like Facebook messenger"* (*# p31*)

Since the study was aimed to assess participants attitudes towards IoT threats, both groups (*Intervention*, *Control*) were asked if they already monitored their devices, and if not, if they would likely start as a result of engaging in this study. If only a small number of participants previously monitored their device activity, this could suggest the risk of threats was not warranted to be sufficient to require device monitoring. If after engaging in the study participants reported they may start monitoring devices, this could suggest their initial lack of awareness contributed to their risk-aversion position. The vast majority of participants in both groups reported they did not previously monitor their smart device activity.

> *"takes too long"* (*# p3*)

> *"never got round to it"* (*# p10*)

It was also very evident that a lack of knowledge and perceived risk of threats contributed significantly to the general lack of device monitoring. Participants reported

being unaware of threats or that their data could be at risk.

> *"I didn't know my devices were at risk" (# p1)*
>
> *"didn't think too much about the risks to be honest" (# p20)*
>
> *"didn't think anyone would be interested in my data" (# p7)*

However, when asked if would they change their behaviour as a result of engaging in this study, a significant number of participants advised they would. It would appear that once they had made aware of the potential threats and the likeliness of their devices being compromised, they considered the risk sufficient enough to start device monitoring.

> *"It's probably something that I can do so should start being more proactive" (# p5)*
>
> *"To be honest I don't think people are aware this is a problem, but it has certainly given me something to think about" (# p9)*
>
> *"risk seems quite real, so yes I should consider starting" (# p22)*

Finally, participants in the *Intervention* group were asked if they would be more likely to monitor their smart device and network activity if they had access to the **Aural**$^{(Au)}$ and **Verbal**$^{(Ve)}$ at home, and if so, would they feel better equipped to detect unusual smart device activity in the future. Participants overwhelmingly reported they would be more likely to monitor devices, and feel equipped to detect unusual activity in the future.

> *"they would give me a better awareness of what was going on my network" (# p3)*
>
> *"it looks quite simple and quick, so I would be more likely to check them" (# p30)*
>
> *"It would be easy to use the Alexa when at home, and the Telegram app when out and about" (# p20)*
>
> *"I think they would encourage me to think more about what my devices are doing" (# p5)*

It was clear from the responses and results in Section 6.3.2 participants found the agents to be an effective way to monitor smart device activity. The feedback and suggested improvements were used to make amendments and refinements to the agents ready for use in the final longitudinal study in the next Chapter.

In this study, a large amount of quantitative data was collected and used to test the hypotheses in Section 6.2.1.

Hypothesis A: Situational Awareness of threats will be improved using conversational agents.

Table 6.5 shows the Pre and Post-Study Cyber Situational Awareness Score(*CSAS*) for both the control and intervention group. The results show that the **Post-Study** *CSAS* score (12.17) for the *Intervention* group was considerably higher than their equivaent **Pre-Study** *CSAS* score (7.23), therefore **Hypothesis H_1** was accepted. Participants reported being more confident at detecting threats when using the conversational agents suggesting situational awareness had been improved. However, it was important to confirm that changes in situational awareness were derived from using the agents, and not from other influencing factors such as the IoT video they watched. To confirm the video had not contributed to confounding the **Pre-Study** *CSAS* score (7.23) for the *Intervention* group was compared with the **Post-Study** *CSAS* score (7.37) for the *Control* group. Since these scores were both derived after watching the IoT video they were considered a fair comparison, and would ensure that the influence of the IoT video in both groups was equal (See Figure 6.1). There was no significant difference found between the two groups therefore **Hypothesis H_3** was accepted.

Finally, results for the *Control* group showed that their **Post-Study** *CSAS* score (7.37) was lower than their equivalent **Pre-Study** *CSAS* score (10.51) therefore **Hypothesis H_2** was accepted. After watching the video explaining IoT threats participants reported being less confident in their ability to detect threats. This result appears to confirm the assertion in [19] that suggested participants may have some awareness of potential threats, but may not fully understand the implications or potential for misuse. Once they had a better understanding of the threats, their reported confidence score more accurately reflected their level of confidence. In this chapter the question "*Are conversational agents a viable method for making users aware of threats in consumer IoT networks?*" was explored. It is clear from the results that the two conversational agents had a positive impact on participants confidence to detect threats. This evidence provides the justification for the further research presented in this thesis, in particular measuring the utility of the agents, and the extent to which performance metrics such as accuracy and efficiency can be improved.

6.5 Conclusions

Previously, Chapter 5 demonstrated how situational awareness of threats could be improved by presenting users with additional information about smart device activity in their environment. The study demonstrated how a **Visual**$^{(Vi)}$ modality could be used effectively to increase awareness of threats. This chapter presented the results of a cross-sectional study which explored the viability of using conversational agents, based on **Aural**$^{(Au)}$ and **Verbal**$^{(Ve)}$ modalities, to improve Cyber Situational Awareness. Participants reported increased confidence in identifying threats when using the two agents developed in this study. The findings of this cross-sectional study now serve as a basis for a final longitudinal study presented in the next chapter, where the use of the two agents will be extensively tested.

Chapter 7

Longitudinal Study to Assess the Utility of Conversational Agents to Improve Cyber Situational Awareness

Chapter 6 presented the cross-sectional study aimed at exploring the viability of conversational agents for improving Cyber Situational Awareness. Two agents based on **Aural**$^{(Au)}$ and **Verbal**$^{(Ve)}$ modalities were created and tested in a cross-sectional study. The results of the study were promising with participants reporting increased confidence in threat awareness when using the agents. The study successfully demonstrated the suitability of conversational agents for aiding improved situational awareness, and provided the justification for further and deeper investigation. Based on the feedback from the cross-sectional study the two conversational agents were further refined. In this chapter, the results of a longitudinal study are presented, where the utility of agents were assessed over a longer period of time.

7.1 Introduction

The focus of this chapter is attempting to answer sub research question SQ4: "*Are conversational agents effective in making users situationally aware of threats in consumer IoT networks?*". Consistent with the methodology used in Chapter 6 a quantitative approach was used to examine the use of the conversational agents for improving Situational Awareness of threats. Mica Endsley's [1] model was again used to assess how

participants perceive device activity, comprehend this in the context of their environment, and use the knowledge to determine if a threat exists. However, since the aim was to fully explore the effectiveness of the agents for improving situational awareness, the three layers of Endsley's model were investigated separately. In addition, as discussed in Section 3.3.4 the International Organisation for Standardisation (ISO) state that systems should be evaluated for usability in terms of a users ability to achieve goals effectively, efficiently and with satisfaction [164]. Therefore, the study was designed to collect data relating to these metrics and also the three individual layers of Endsley's model namely: **Perception**, **Comprehension** and **Projection**.

Finally, a qualitative approach was also taken to examine the usability and utility of the conversation agents. Following completion of the study presented in this chapter, short structured interviews were undertaken to elicit feedback, and assess if participants reported increase in confidence and awareness of threats matched the empirical evidence collected.

7.2 Methodology

7.2.1 Experimental Variables

This study explored the usability and utility of the conversational agents in depth. Previously in Chapter 6, the three levels of Endsley's model were combined to create a single metric (CSAS). In this study, the aim was to explore each level separately, therefore, the mean **Perception** ($\overline{pe}$), **Comprehension** ($\overline{co}$), and **Projection**($\overline{pr}$) were calculated individually using the statements created in Section 3.3.5. Data collected was ordinal and used as a dependent variable. In addition, the usability of each agent was measured. This involved measuring the *efficiency* (in seconds) with which participants could assimilate information about events in their environment, synthesise this into a meaningful understanding of the situation, and the *accuracy* with which they could identify threats in a network. Finally, participants satisfaction when using the agents was measured. For this study, accuracy, efficiency and satisfaction were used as dependent variables. While the **Aural**$^{(\mathrm{Au})}$ and **Verbal**$^{(\mathrm{Ve})}$ conversational agents were used as independent variables. For consistency, a Pre-Study/Post-Study design was used for analysis of the dependent variables and the reported differences in Cyber Situational Awareness ($\overline{pe}$, $\overline{co}$, $\overline{pr}$), accuracy, efficiency and satisfaction was measured. The study hypotheses are therefore defined as:

Hypothesis A: Situational Awareness of threats will be improved using conversational agents.

The study in chapter 6 demonstrated the suitability of conversational agents for aiding

improved situational awareness. Overall participants reported increased confidence, however, the aim of this study was to explore if the increased confidence was consistent across the three levels of Endsley's SA model [1]. To test this assumption it was hypothesised that:

> **H_1:** Post-Study Mean **Perception** ($\overline{pe}$) score will be higher than the Pre-Study Mean **Perception** ($\overline{pe}$) score.
>
> **H_2:** Post-Study Mean **Comprehension** ($\overline{co}$) score will be higher than the Pre-Study Mean **Comprehension** ($\overline{co}$) score.
>
> **H_3:** Post-Study Mean **Projection** ($\overline{pr}$) score will be higher than the Pre-Study Mean **Projection** ($\overline{pr}$) score.
>
> **H_4:** Post-Study **Perception** ($\overline{pe}$), **Comprehension** ($\overline{co}$) and **Projection** ($\overline{pr}$) scores of the ***Control*** group will **Not** be significantly different than the Pre-Study $\overline{pe}$, $\overline{co}$ and $\overline{pr}$ scores.

Hypothesis B: Performance of detecting threats will be improved using conversational agents.

In addition to exploring if participants reported feeling more confident when using conversational agents, this study also assessed if the use of the agents would improve participants performance when detecting threats. To test this assumption it was hypothesised that:

> **H_5:** Mean Detection **Efficiency** (in seconds) will be lower using a conversational agent than when using the baseline visual method.
>
> **H_6:** Mean Detection **Accuracy** will be higher using a conversational agent than when using the baseline visual method.

7.2.2 Study Design

A longitudinal study was conducted to evaluate the utility of using conversational agents to improve Cyber Situational Awareness. Previously, in Chapter 6 the cross-sectional study collected data from a large population of users at a single point in time. In this study, the aim was to collect data from a smaller sample of users over an extended period [149] lasting twenty-one days (See Figure 7.1). For consistency, the study was similar in design to Chapter 6 (See Figure 7.2) where participants were recruited and again assigned to two groups as described in Section 7.2.4. Participants in the *intervention* group watched an introductory video about threats facing smart homes, and completed a short related activity. The group were then asked to complete a Pre-study survey

and indicate their level of agreement with nine statements (See Appendix I) relating to their awareness and ability to monitor smart device and network activity. These were mapped to Mica Endsley's SA model as described in Section 3.3.5. Participants were supplied with an Amazon Echo device (preconfigured with the **Aural**$^{(Au)}$ agent), and had the Telegram app with the **Verbal**$^{(Ve)}$ agent setup on their supplied mobile device. For the remainder of the four day Pre-Study period participants familiarised themselves with the two conversational agents and a visualisation tool (web application), used as a baseline to compare the conversational agents against (See Section 7.2.2). It was important to have a baseline comparison since in the cross-sectional study in Chapter 6, a significant number of participants reported not monitoring their smart devices or network prior to completing the study. It was, therefore, reasonable to expect that this may be true of participants in this study, in which case some degree of improvement from using the agents would be expected. It was equally important to recognise that some participants did report using monitoring software previously, therefore, again it was reasonable to expect that this may also be true for some participants in this study. Since the aim in this study was to quantify the degree of improvement gained from using the conversational agents, observed improvements were measured against the baseline visualisation tool. To avoid introducing bias for participants who may have previously used a monitoring tool and could be familiar with a particular layout/representation, the decision was made not to use an existing commercial or open source visualisation tool (e.g. mcafee, wireshark, splunk), but rather develop a simple bespoke tool for use in this study (See Section 7.2.2).

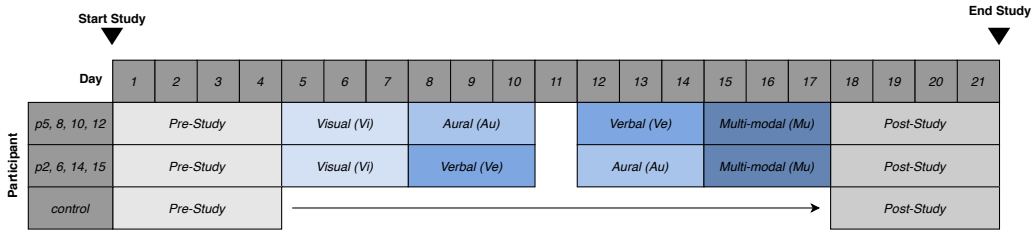


Figure 7.1: Longitudinal Study Timeline

The main section of the study lasted twelve days and was split into four sub-studies: **Visual**$^{(Vi)}$, **Aural**$^{(Au)}$, **Verbal**$^{(Ve)}$ and **Multi-modal**$^{(Mu)}$ (See Figure 7.1). Each sub-study lasted three days and participants used the associated conversational agent or visualisation tool daily to answer questions relating to use-case scenarios presented in Appendix D.2 - D.5. In the fourth sub-study (**Muli-modal**$^{(Mu)}$) participants were able to use any combination of tools to answer the daily questions. To further avoid bias, the *intervention* group was split in half with four participants (*p5,8,10,12*) randomly assigned to use the **Aural**$^{(Au)}$ agent in study two and **Verbal**$^{(Ve)}$ agent in study three,

and the remaining participants (*p2,6,14,15*) assigned to use the agents in reverse order. During the undertaking of the four sub-studies participants answers to the use-case scenario questions were measured for accuracy, efficiency and usability. Finally, during the four day Post-Study period, participants completed a Post-study survey of the same nine statements, and variance in their attitudes was recorded. Post-Study interviews were also conducted with the *Intervention* group during this period, to elicit feedback about the study and use of the conversational agents.

The use of a *control* group again enabled confounding variables to be mitigated and scientific rigour enhanced. Participants in the *control* group watched the introductory video about threats facing smart homes, completed the related activity, before completing the same Pre-Study survey. Participants did not then engage with the study between *day 5* and *day 18*, before finally completing the Post-Study survey. This approach was used to check that variables outside of the study had not influenced the results in the *intervention* group. If the *control* group situational awareness scores remained roughly the same Pre-Study/Post-Study, the validity of the data could be confirmed, since any variance in situational awareness scores in the *intervention* group would be derived from using the conversational agents.

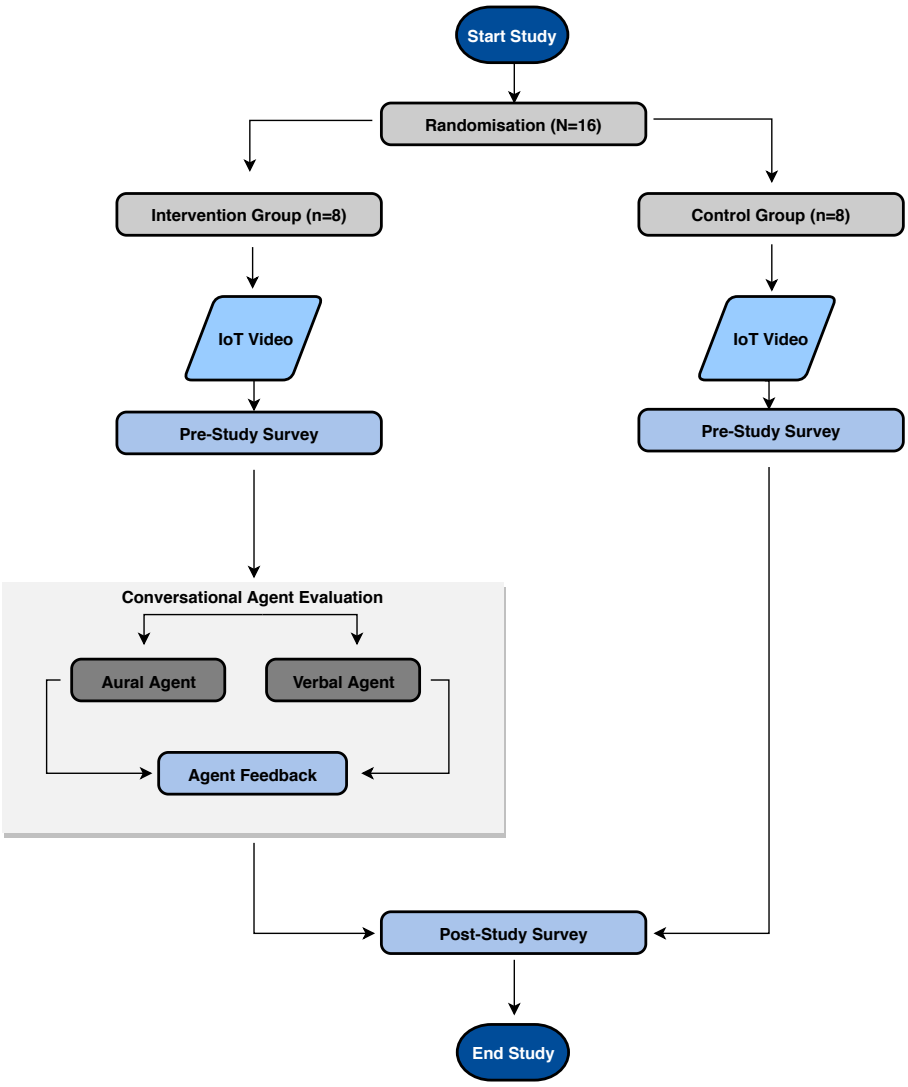


Figure 7.2: Longitudinal Study Design

Use-Case Scenario Design (Amendments from Chapter 6)

For consistency, the seven use-cases from the previous study in Chapter 6 were adopted for use in this study. In addition, a further two use-cases were added as detailed in Table 3.2. The corresponding scenario devised for each use-case is presented in Appendix D.2 - D.5. As discussed in Section 3.3.2, since the study used the same nine use-cases for each of the four studies, it was important to have different scenarios to represent each use-case. This was again important to avoid confounding variables or adding bias into the research. If a participant had seen a scenario in a previous study it may effect their decision making or performance. By using different scenarios to represent the same use-case, confidence could be attained in the reliability of the collected data. Table 3.3 demonstrates how each scenario maps to the corresponding use-case in the study.

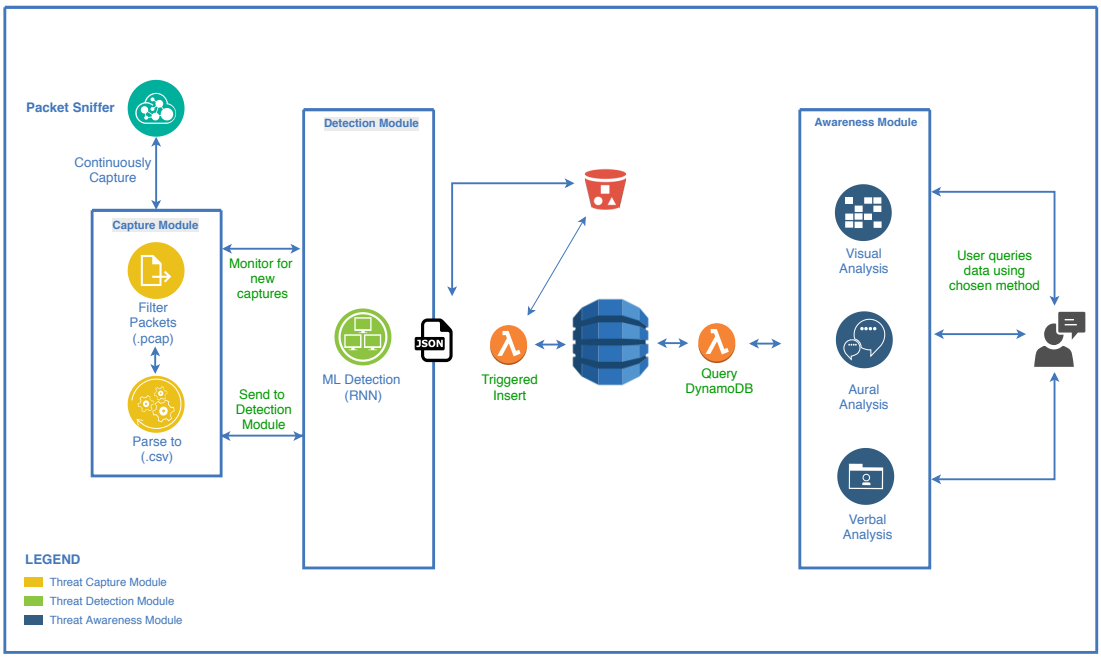


Figure 7.3: Conversational Cyber Situational Awareness Framework

Cyber Situational Awareness was measured using the nine confidence statements created in Section 3.3.5. A survey was generated which included the nine confidence statements, plus seven additional statements relating to general smart device security (See Table 6.1). The additional general security questions were again only used as distraction [200] to disguise the focus of the nine confidence statements. As previously discussed, this was important as previous studies have emphasised the importance of avoiding the introduction of bias into a study by priming participants [201, 202]. The use of the seven additional statements ensured participants were unaware that the focus of the study was on their confidence to detect threats, which could have resulted in inflated confidence levels being reported. For each statement in the survey participants were asked to indicate their level of agreement using a five-point likert scale from *Strongly Disagree (1)* to *Strongly Agree (5)*. Finally, participants completed the same survey before and after using the conversational agents, and the difference in their responses was compared. The Pre-Study/Post-Study Survey is presented in Appendix I.

Figure 7.3 shows the conceptual Conversational Cyber Situational Awareness Framework presented in this thesis. The framework is made up of three modules: *Capture module, Detection module* and *Awareness module*. The *Capture* module is the **ETL Pipeline** presented in Section 6.2.3 The *Detection* module is the **BLSTM-RNN** IDS

presented in Section 4.2.5. Finally, the *Awareness* module is the Conversational Agents developed in Section 6.2.3 and 7.2.3. The *Awareness* module also contained the baseline Visualisation tool which was developed for this study and is presented in Figure 7.4. The tool was developed using the *Gentelella Bootstrap admin template*[1] and was designed to have a similar look and feel to standard dashboards a participant may have used previously in other contexts.

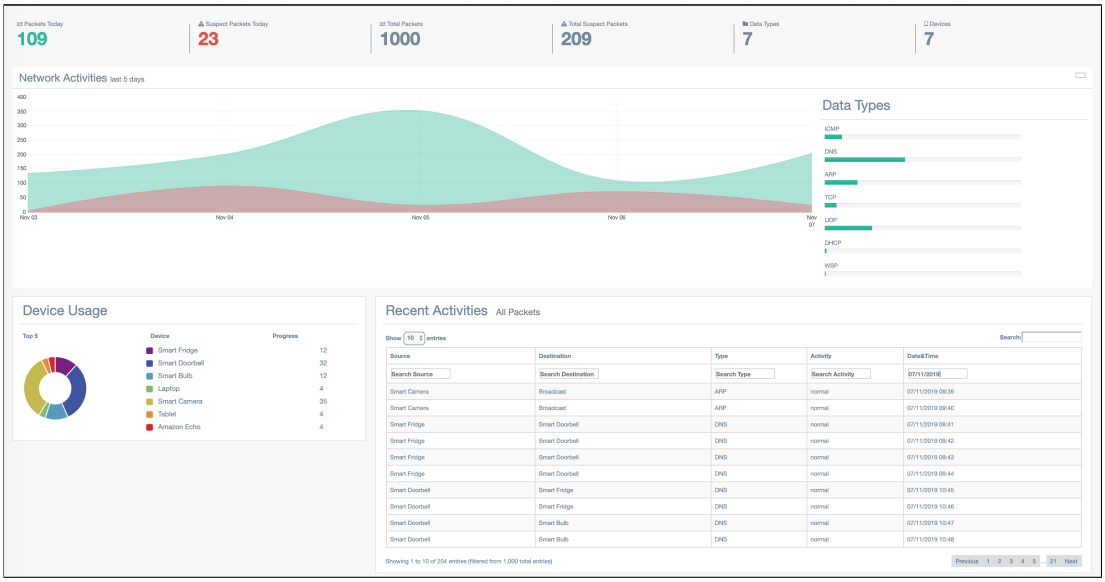
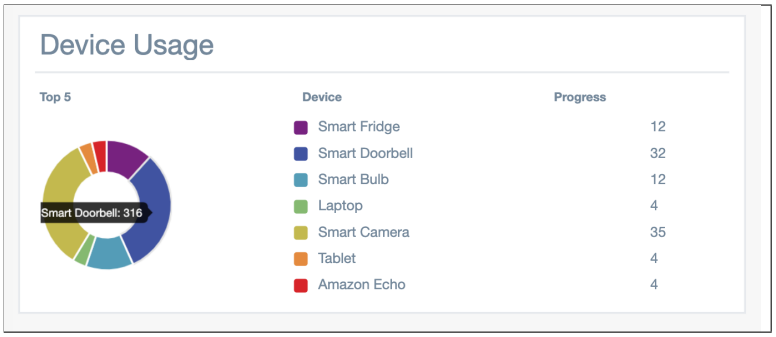
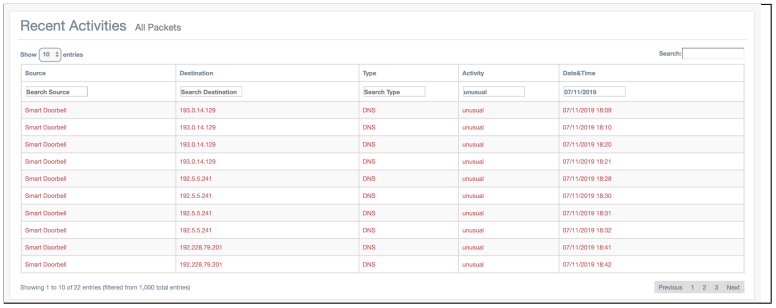


Figure 7.4: Baseline Visualisation Tool

[1] https://colorlib.com/polygon/gentelella/



(**a**) Device Usage



(**b**) Recent Activities

Figure 7.5: Baseline Visualisation Tool Elements

From the dashboard participants were able to find information about network activity for a five day period. Participants could hover over elements of the dashboard to quickly see individual device information such as device usage as shown in Figure 7.5a. Participants could also quickly search for specific features of activities using the search facility as shown in Figure 7.5b.

Hawthorne Effect

This study explored the responses from two groups (*Control* and *Intervention*). For the intervention group consideration was given to the impact of the Hawthorne Effect on the study. The Hawthorne Effect concerns research participation, the consequent awareness of being studied and possible impact on behaviour [162]. It has been reported that participants may behave differently in lab-based experiments due to being in a different environment, and the knowledge of being observed [208] [209]. To mitigate the potential of this phenomenon the study was designed to use Naturalistic observation, where participants completed the study at home rather than in a lab environment. The goal was to observe their behaviour from a distance, in a natural setting without intervention [210]. By doing so, this approach did introduce a risk that the results may be less accurate. For example, if a participant was distracted while completing a

task, the time taken to complete the task could be increased. However, the study was designed to mitigate this by including a pause break between each task, to allow for such distractions and ensure time duration was only recorded once a participant had started a task.

7.2.3 Conversational Agent Design (Amendments from Chapter 6)

The conversational agents remained largely the same as presented in Section 6.2.3. Minor bug fixes were addressed and performance improvements were made in the backend AWS infrastructure. Participant feedback from the cross-sectional study in Chapter 6, provided suggestions of how the agents could be improved:

> *"option to see which devices have been used most each day" (# p3)*
>
> *"add more questions I can ask Alexa" (# p2)*
>
> *"quick summary for multiple days" (# p10)*
>
> *"add the option of a weekend summary" (# p16)*
>
> *"multiple day summary" (# p37)*

In response to this participant feedback, two additional intents were added to both the **Aural** and **Verbal** agents as described in the next section.

Agent Intents (Amendments from Chapter 6)

For consistency, the ten *intents* from the previous study in Chapter 6 were adopted for use in this study. In addition, in response to user feedback from the previous study, a further two intents were added to this study as detailed below:

11. ***unusualActivityLastThreeDays:*** Responds to a user query and returns a summary of normal and unusual activity on each of the last three days.
12. ***mostActiveSrcDevLastThreeDays:*** Responds to a user query and returns a list of three most active source devices on each of the last three days.

As a result, a total of twelve intents were created and used within this study. During the main section of the study, which lasted twelve days, participants were able to choose which intents they would use to query the agents for information. The participants then used the responses to their queries to answer the daily questions relating to use-case scenarios presented in Appendix D.2 - D.5. Table 7.1 shows how each intent mapped to the nine situational awareness statements which participants completed in the Pre-Study/Post-Study surveys.

Table 7.1: Situational Awareness and Intent Mapping

Situational Awareness \ Agent Intents	*i6*	*i1*	*i11*	*i2*	*i3*	*i4*	*i5*	*i7*	*i8*	*i12*	*i9*	*i10*
pe1				x	x	x	x	x	x	x		
pe2					x			x	x	x	x	
pe3				x	x			x	x	x	x	
co1	x	x	x	x		x	x					x
co2					x	x	x	x	x	x	x	
co3					x	x	x	x	x	x	x	
pr1	x	x	x	x		x	x					x
pr2	x	x	x	x	x	x	x				x	
pr3			x			x	x	x	x	x	x	x

7.2.4 Participants

In this study, the aim was to assess the utility of conversational agents for improving awareness of threats facing smart devices. Sixteen participants were recruited for this final study, which took place between September and October 2019. The study was advertised within the university and through LinkedIn and Social Media. This approach enabled a wide range of views to be collected and oversampling of a specific demographic to be avoided, namely the local student population. Interested parties needed to meet the following criteria, to be eligible to participate:

- Be over 18 years of age and be able to provide informed consent;
- Own or have previously used a smart device;
- Be available for the full 21 day study;
- Not completed any previous studies associated with this research.

Participants provided informed consent by reading the study agreement (See Appendix F), before indicating their consent to participate when clicking to proceed to the next page of the Pre-Study survey.

Additional Participant Profile Screening

Further checks were completed on the suitability of each participant during the Pre-Study survey. Participants were asked if they had any medically diagnosed visual, auditory or learning difficulties. These qualities were requested to allow an assessment

to be made whether such difficulties could affect the participants performance and the validity of the collected data.

- **Participant 8:** reported a 30% reduction in hearing, however, confirmed that daily activities were not affected.
- **Participant 2:** reported having Astigmatism, however, confirmed that this was corrected using contact lenses so does not affect daily activities.
- **Participant 14** reported having dyslexia and issues with grammar and spelling.

Following discussions with each participant it was decided that the reported qualities did not pose a risk to the validity of the study.

Participant Demographics

Participant demographics are presented in Table 7.2, with the largest demographic being Intermediate users aged 25-39. All participants reported owning or having previously used smart devices.

Table 7.2: Participant Demographic

Gender	%	***Age***		%	**Ability**		%
Male	8 (50)	**18-24**	4	(25)	**Novice**	3	(19)
Female	8 (50)	**25-39**	7	(44)	**Intermediate**	7	(44)
		40-59	3	(19)	**Advanced**	5	(31)
		60+	2	(12)	**Expert**	1	(6)

n = 16

In accordance with recommendations for well-designed experiments [157], participants were randomly assigned to the two groups (Control and Intervention) as shown in Table 7.3.

Table 7.3: Group Allocation

	Novice	**Intermediate**	**Advanced**	**Expert**
Control	**1** (p9)	**3** (p1, p3, p7)	**3** (p11, p13, p16)	**1** (p4)
Intervention	**2** (p2, p12)	**4** (p6, p8, p10, p15)	**2** (p5, p14)	**0**

All sixteen participants completed the study, therefore, attrition was not present in the study. In addition, since the study design followed guidelines for well-designed experiments [157], the intervention results presented in this thesis are believed to have validity.

7.3 Results

This section presents the results of the study using the metrics stated in Section 7.2.1, namely usability and situational awareness. For usability, the study measured participants accuracy and efficiency of detecting threats, and also their satisfaction of using the conversational agents. For situational awareness, the study assessed participants ability to be aware of threats in their surroundings by considering how they perceive, understand and react to situations.

7.3.1 Usability

Detection Accuracy

The first metric measured how accurately participants could detect threats using the conversational agents. Specifically, it recorded their accuracy of collecting the necessary information about smart device activity, and how well this was used to determine if threats had occurred. Table 7.4 shows the Precision (P), Recall (R) and F-measure (F_1) scores for each of the nine use-cases described in Section 7.2.2. The Mean, Median and Standard Deviation for the agents is presented in Table 7.5 which shows an improvement in mean precision (1.0) when using the **Verbal**$^{(Ve)}$ agent compared with the respective score (0.973) for the baseline **Visual**$^{(Vi)}$ method, suggesting false positive detections were present when using the visual method. Recall scores show that false negative detections (0.987) were also present when using the baseline visual method, but not the verbal agent. Precision (0.869), Recall (0.924) and F-measure (0.887) were noticeably lower when using the **Aural**$^{(Au)}$ agent compared with the baseline visual method. Analysis of the data shows that false positive detections in particular contributed to the lower f-measure score (0.887) demonstrating that in some cases the participants had gathered the necessary information but had failed to comprehend its meaning, resulting in incorrect answers provided. Finally, when participants chose to use a **Multi-modal**$^{(Mu)}$ approach using a combination of agents and baseline visual method, perfect precision and recall was achieved.

Table 7.4: Precision, Recall and F-Measure Detection Accuracy of Threats

		uc1	uc2	uc3	uc4	uc5	uc6	uc7	uc8	uc9
Visual$^{(Vi)}$	**Precision**	1.0	1.0	1.0	1.0	1.0	1.0	0.88	0.88	1.0
	Recall	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.88
	F-Measure	1.0	1.0	1.0	1.0	1.0	1.0	0.93	0.93	0.93
Verbal$^{(Ve)}$	**Precision**	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Recall	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	F-Measure	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Aural$^{(Au)}$	**Precision**	1.0	1.0	0.86	1.0	0.83	1.0	0.75	0.63	0.75
	Recall	1.0	1.0	0.86	1.0	0.71	0.75	1.0	1.0	1.0
	F-Measure	1.0	1.0	0.86	1.0	0.77	0.86	0.86	0.77	0.86
Multi$^{(Mu)}$	**Precision**	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Recall	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	F-Measure	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 7.5: Mean, Median and Standard Deviation: Precision, Recall and F-Measure

	Precision (P)			**Recall (R)**			**F-Measure (F_1)**		
	$\overline{x}$	$\tilde{x}$	σ	$\overline{x}$	$\tilde{x}$	σ	$\overline{x}$	$\tilde{x}$	σ
Vi	0.973	1.0	0.53	0.987	1.0	0.04	0.977	1.0	0.035
Ve	1.0	1.0	0.00	1.0	1.0	0.00	1.0	1.0	0.00
Au	0.869	0.86	0.14	0.924	1.0	0.12	0.887	0.86	0.093
Mu	1.0	1.0	0.00	1.0	1.0	0.00	1.0	1.0	0.00

$n = 8$

Detection Efficiency

The second metric measured how efficiently participants could detect threats using the conversational agents. Specifically, it recorded how long (in seconds) it took to collect the necessary data about smart device activity, process this into meaningful information which they could use to determine if a threat had occurred. Table 7.6 shows recorded detection times for each of the nine use-cases described in Section 7.2.2. The Mean, Median and Standard Deviation for each agent is presented in Table 7.7 which shows that participants were less efficient when using the **Verbal**$^{(Ve)}$ (473.13) and **Aural**$^{(Au)}$ (539.52) agents compared to when using the baseline **Visual**$^{(Vi)}$ (460.49) method.

However, when participants were free to chose a **Multi-modal**$^{(Mu)}$ approach using a combination of agents and baseline visual method, efficiency was improved (455.17).

Table 7.6: Detection Efficiency

	uc1	**uc2**	**uc3**	**uc4**	**uc5**	**uc6**	**uc7**	**uc8**	**uc9**
Vi	26.25	91.12	30.33	27.13	59.90	27.33	60.64	55.38	82.42
Ve	35.73	41.62	65.54	53.88	64.19	54.65	58.41	47.58	51.53
Au	33.16	50.91	74.55	43.29	73.99	64.38	74.26	66.34	58.63
Mu	33.29	56.06	43.83	44.20	66.20	41.39	57.88	49.88	62.45

efficiency shown in seconds (s)

Table 7.7: Mean, Median and Standard Deviation: Detection Efficiency

	$\overline{x}$	$\tilde{x}$	σ
Vi	460.49	443.94	54.02
Ve	473.13	478.20	12.62
Au	539.52	540.31	11.93
Mu	455.17	459.51	21.70

$n = 8$

To check if differences were significant a Friedman test was carried out to compare participant efficiency scores for the four detection methods (Vi, Ve, Au, Mu). The test showed there was a statistically significant difference in efficiency between the four methods, $\chi^2(3) = 12.900$, $p = .005$

Satisfaction: System Usability Questionnaire

The usability of any system has to be viewed in terms of the context in which it is used, and its appropriateness to that context [164]. This view is reflected in the ISO 9241-11 standard which provides a framework for understanding the concept of usability and its application to interactive systems. In compliance with this standard, the study sought to asses user satisfaction relating to the two conversational agents. In Table 7.8 the system usability scores for each conversational agent are presented. Scores for each participant were calculated using the methodology described in Section 3.3.4. As described in [164] SUS scores are presented on a scale from 0 to 100. The final SUS score was calculated by first determining the sum of each item. The score contribution for odd numbered questions was adjusted to be the scale position minus 1. The scale

contribution for even numbered questions was adjusted to be 5 minus the scale position. Finally, the sum of the scores was multiplied by 2.5 to obtain the overall value of system usability.

Table 7.8: System Usability Scale (SUS) Scores

	p2	*p5*	*p6*	*p8*	*p10*	*p12*	*p14*	*p15*
Ve	75	80	52.5	65	75	70	80	95
Au	52.5	47.5	40	70	80	60	95	87.5

To interpret the results, a score of 68 was used as a measure of usability. A score falling close to this point can be assumed to have average usability [164]. The tests conducted in this study found that six participants scored the **Verbal**$^{(Ve)}$ agent above average, compared to only four participants scoring above average for the **Aural**$^{(Au)}$ agent. These results suggest that both agents require further improvement.

The validity of the study data was tested using techniques consistent with [211, 212, 213]. A Cronbach alpha coefficient test was therefore performed for the SUS results of both conversational agents. The test checked for internal consistency, where a reliable scale is said to have a Cronbach's alpha coefficient above .70 [207]. Cronbach alpha coefficients for the **Verbal**$^{(Ve)}$ and **Aural**$^{(Au)}$ agents exceeded this recommended value with $\alpha = .89, .93$ respectively, therefore the data was considered valid with good internal consistency

7.3.2 Cyber Situational Awareness

The study measured participants ability to be situationally aware of threats in their environment using Endsley's reference model [1]. To achieve a level of awareness in any given situation the three levels of the model, namely Perception (Pe), Comprehension (Co) and Projection (Pr), are combined. To measure the effectiveness of the conversational agents to improve situational awareness, participants were asked to indicate their agreement with the SA statements in Table 6.1 before and after using the agents to answer questions in each of the four studies (See Appendix H). Table 7.9 shows Pre and Post-Study situational awareness scores. The Mean, Median and Standard Deviation for each of the three elements in Endsley's SA model is presented in Table 7.10. The results show that mean Post-Study perception (4.13), comprehension (3.92) and projection (3.75) scores improved as a result of using the conversational agents.

Table 7.9: Pre-Study Post-Study Situational Awareness Scores

		Perception			Comprehension			Projection		
	#	pe1	pe2	pe3	co1	co2	co3	pr1	pr2	pr3
Pre-Study	***p2***	4	2	1	2	2	1	1	2	3
	p5	2	2	3	3	3	3	3	3	3
	p6	3	2	2	3	3	3	2	2	2
	p8	2	2	2	3	3	2	2	2	2
	p10	4	3	2	2	2	3	3	3	2
	p12	4	2	1	2	2	2	2	2	2
	p14	4	4	3	3	3	2	3	3	3
	p15	2	2	1	3	2	1	2	3	2
Post-Study	***p2***	4	3	3	4	4	3	4	4	3
	p5	4	4	4	5	4	4	4	4	4
	p6	4	4	4	4	4	3	3	3	3
	p8	5	4	4	4	4	4	4	4	4
	p10	4	4	4	4	4	3	4	4	4
	p12	5	4	4	5	5	4	4	4	4
	p14	5	5	5	4	4	4	4	4	5
	p15	4	4	4	4	4	2	3	3	3

To check if differences between the *Intervention* group Pre-Study/Post-Study scores were significant, Wilcoxon Signed-Rank tests were carried out to compare the three elements of situational awareness (Perception, Comprehension and Projection). The Post-Study **Perception** median score (4.00) was found to be higher than the Pre-Study score (2.33). A Wilcoxon Signed-Rank test indicated that the difference was statistically significant, Z= -2.53, p= *.011*. The Post-Study **Comprehension** median score (3.83) was found to be higher than the Pre-Study score (2.50). A Wilcoxon Signed-Rank test indicated that the difference was again statistically significant, Z= -2.585, p= *.010*. Finally, the Post-Study **Projection** median score (4.00) was also found to be higher than the Pre-Study score (2.17). A Wilcoxon Signed-Rank test indicated that the difference was again statistically significant, Z= -2.530, p= *.011*.

The Pre-Study/Post-Study Mean, Median and Standard Deviation for the *Control* group are presented in Table 7.10. **Perception** scores were found to be consistent between studies, however, a small variance in **Comprehension** and **Projection** were recorded. To check if variances were significant a Wilcoxon Signed-Rank test was again carried out which showed the variances in **Comprehension** and **Projection** were not

statistically significant, Z= -.324, p= *.746*, Z= -.577, p= *.564* respectively.

Table 7.10: Mean, Median and Standard Deviation: Cyber Situational Awareness

		Perception			**Comprehension**			**Projection**		
		$\overline{x}$	$\tilde{x}$	σ	$\overline{x}$	$\tilde{x}$	σ	$\overline{x}$	$\tilde{x}$	σ
CTL	**Pre-Study**	2.75	2.67	.79	2.79	2.67	.73	2.25	2.17	.77
	Post-Study	2.75	2.33	.79	2.71	2.50	.97	2.29	2.00	.72
INT	**Pre-Study**	2.46	2.33	.62	2.42	2.50	.05	2.38	2.17	.45
	Post-Study	4.13	4.00	.47	3.92	3.83	.43	3.75	4.00	.05

To test the validity of the data a Cronbach alpha coefficient test was conducted on both groups Pre and Post-Study situational awareness scores. Cronbach alpha coefficient for the *Control* and *Intervention* groups (α = .93, .96 respectively) were found to be above the recommended value .70 [207], therefore, the data was considered valid with good internal consistency.

As a final measure of validity, the equation in Section 6.2.1 was used to calculate and compare the Pre-Study *CSAS* score for the *Intervention* group (7.26) in this study with the corresponding *CSAS* score for the *Intervention* group in the cross-sectional study (7.23) (See table 6.5). The *CSAS* scores were found to be consistent, providing further validity to the data.

7.4 Discussion

This study was undertaken to assess the utility of conversational agents for improving Cyber Situational Awareness. Specifically, the aim was to explore if user awareness and perception of threats facing consumer IoT networks could be improved through the use of the agents. The mean **Perception** ($\overline{pe}$), **Comprehension** ($\overline{co}$) and **Projection**($\overline{pr}$) was calculated and used as metrics to measure if participants Cyber Situtaional Awareness improved Pre-Study/Post-Study. In addition, the aim was also to assess the usability of the agents. Participants in the *Intervention* group were asked to use the two conversational agents to answer questions relating to use-case scenarios presented in Section 7.2.2. To measure the usability of the agents the study calculated the *efficiency* (in seconds) with which participants could assimilate information about events in their environment, synthesise this into a meaningful understanding of the situation, and the *accuracy* with which they could identify threats in a network. To measure how satisfied participants were when using the agents, a SUS score for each agent was calculated. Finally, during the four day Post-Study period (See Figure 7.1)

short interviews were conducted with the *intervention* group to elicit feedback about the study and agent use (See Appendix J). For brevity, a representative sample of responses are provided below for each of the questions asked. First, participants were asked to provide feedback about their overall experience of using the conversational agents. Participants all reported a positive experience, stating they found both agents to be user friendly and easy to use.

> *"Overall my experience using these devices [conversational agents] was quite positive. I found them quite exciting and interesting to use. Simple to use and quite clear in the responses [answers to queries] "* (*# p8*)

> *"I found both the conversational agents user friendly and fairly easy to use once I had a play around with them. "* (*# p10*)

Next, participants were asked about their experience of using both agents to answer questions relating to the three areas of investigation: **Perception**, **Comprehension** and **Projection**. Specifically, their experience of using the agents to find out which devices were using the network and how much activity they had (**Perception**). Their experience of using the agents to work out if a smart device or network was functionally normally (**Comprehension**). And finally, their experience of using the agents to determine if a smart device had been compromised, or an attack had taken place on the network (**Projection**). For each of the three areas, participants were also asked if they preferred using one of the agents to answer the question or the baseline visualisation tool. For queries (*intents*) which elicited quick or short summary responses, participants reported finding the conversational agents more convenient than the visualisation tool. This was evident for intents such as *"what is the status of the network today ?"* (*i6*) or *"give me a summary of today's activity"* (*i1*).

> *"In this case when it's just a simple question of "is there something unusual happening or not" then probably the conversational agents were better because it was quick and simple to ask, and in the case of someone who is not as computer oriented, it's a way for them to know without having to go into all the nooks and crannies of the visualisation tool."* (*# p14*)

> *"For quick check of devices the conversational agents were more convenient. You could get the information from the visual tool but I felt it took longer, plus it was less convenient"* (*# p10*)

For queries (*intents*) which elicited longer responses or more detailed information, participants still enjoyed using the conversational agents, however, found the **Verbal**$^{(Ve)}$

agent to be easier and more efficient. They reported difficulty understanding percentages when using the **Aural**$^{(Au)}$ agent, often requiring the agent to repeat a response several times. This was evident for intents such as *"give me a summary of activity by Smart Camera on 20th January 2019"* (*i3*).

> *"In terms of the Alexa again it's a quick way of asking for the direct information, but I found listening and drawing out percentages [specific detailed information] quite tricky in comparison to telegram, which I found easier to read the information and gather the required information "* (*# p12*)

> *"In general with both Alexa and Telegram, when it comes to finding out which source devices [were using the network] or whether there was anything unusual or normal [activities] they were about the same for me, I had no issues. But, when it came to asking about anything where the answer involved percentages of whether [activities were]normal or unusual, Alexa I struggled with more because I had to repeat [answers] a couple of times, but eventually I did get there. So in terms of that I would prefer the telegram overall "* (*# p14*)

For queries (*intents*) which elicited specific information in a response or which contained long detailed information, participants reported a preference for combing an agent(s) with the visualisation tool to gather the necessary information in a staged approach. This multi-modal preference was reflected in the accuracy and efficient results in Table 7.4 and 7.7, which showed participants were most accurate and efficient when using a multi-modal approach. This was evident for intents such as *"give me activity totals for the Smart Camera for the last three days "* (*i9*) or *"give me details of the first unusual activity on 20th January 2019"* (*i3*).

> *"I would use a combination of all three. I would initially use Alexa or Telegram to see if any issues had been detected, but would then use the visualisation tool to get the detailed information"* (*# p2*)

> *"The conversational agents were good to give me the initial status of the network. This made me want to then use the web interface [visualisation tool] to determine what had actually happened"* (*# p5*)

> *"I would prefer to use the Alexa to determine whether a system or a smart device was functioning normally. But beyond this if there's any unusual activity I would then prefer to use the visualisation tool to narrow down the areas that have been affected"* (*# p8*)

The study next assessed if the participants felt the agents had improved their awareness of threats, understanding of suspicious activity, and their ability to identify a compromised device in the future. If participants reported improvements in these areas, this would suggest that the agents had improved Cyber Situational Awareness, by deriving benefits at all three layers of Endsley's SA model [1]. Participants reported improvements in all areas.

> *"For sure I could tell which devices were using the network and how much and if anything unusual was going on. Before [doing this study] I didn't even think about it or have a clue what they were doing" (# p10)*

> *"Yes because before a device may have looked ok but actually it had been tampered with. Now at least I would have more of an idea if things weren't right " (# p2)*

> *"I believe the conversational agents help me understand more because I didn't know what was going on in the network before" (# p5)*

In the Post-Study survey, participants were asked if they felt it was likely/unlikely that smart devices would be compromised and used as described in this study. They were also asked if they felt the risk of smart devices being compromised is high enough to justify the effort required to monitor them. The participants were asked to expand on the responses they gave to these questions in the survey. This was important in order to gain a deeper understanding of their general awareness and perception of threats facing consumer IoT networks. Participants reported being aware of previous threats, in particular threats targeting smart devices.

> *"I've read in recent articles that the Amazon ring doorbell has been compromised quite severely leading to attacks. But, there are also other ways of getting into a network" (# p5)*

> *"I've read a lot of articles recently about data being stolen, and a lot of the problems have been these home security devices. I feel the risk is therefore high enough to warrant checking the network for these kinds of intrusions" (# p5)*

> *"I feel that, and this is probably very naive of me, that I am not somebody that would be targeted because I am just average Joe Bloggs off the street. I think that's probably why I'm not taking these threats seriously enough [at present] whereas I probably should be" (# p12)*

> *"When I was looking at the information I knew something wasn't right but*

I didn't know what. I've seen stories before on social media Facebook about how the smart devices had been hacked " (*# p2*)

Finally, participants were asked if they would be more likely to monitor their smart device and network activity if they had access to the **Aural**$^{(Au)}$ and **Verbal**$^{(Ve)}$ agents at home, and if so, would they feel better equipped to detect unusual smart device activity in the future. Participants reported they would be more likely to monitor devices, and feel equipped to detect unusual activity in the future.

"I think I probably would monitor smart devices more at home, just because with Alexa and Telegram it's a case of ease of use. I don't need to hook up my laptop or use Wireshark or stuff that I'm normally used to doing, which takes more time to analyse. This [conversational agents] will give me a quick response so makes it easy to check more often than probably I would normally do" (*# p14*)

"Yeah I think for the time it takes to actually monitor the activity I would definitely make an effort to use these facilities [conversational agents]" (*# p8*)

It was clear from the responses in the Post-Study interviews participants found the agents to be an effective way to monitor smart device activity.

In this study, a large amount of quantitative data was collected and used to test the hypotheses in Section 7.2.1.

Hypothesis A: Situational Awareness of threats will be improved using conversational agents.

Table 7.10 shows the Pre and Post-Study Perception ($\overline{pe}$) , Comprehension ($\overline{co}$) and Projection ($\overline{pr}$) for both the *Control* and *Intervention* group. The Post-Study Mean **Perception** ($\overline{pe}$) score (4.13) was found to be higher than the Pre-Study Mean **Perception** ($\overline{pe}$) score (2.46) therefore **Hypothesis H_1** in Section 7.2.1 was accepted.

The Post-Study Mean **Comprehension** ($\overline{co}$) score (3.92) was found to be higher than the Pre-Study Mean **Comprehension** ($\overline{co}$) score (2.42) therefore **Hypothesis H_2** was accepted.

The Post-Study Mean **Projection** ($\overline{pr}$) score (3.75) was found to be higher than the Pre-Study Mean **Projection** ($\overline{pr}$) score (2.38) therefore **Hypothesis H_3** was accepted.

The Post-Study Mean **Perception** ($\overline{pe}$) score (2.75), **Comprehension** ($\overline{co}$) score (2.71) and **Projection** ($\overline{pr}$) score (2.29) were found not to be significantly different than the

Pre-Study $\overline{pe}$ (2.75), $\overline{co}$ (2.79) and $\overline{pr}$ (2.25) scores, therefore **Hypothesis $\mathbf{H_4}$** was accepted.

The results lead to a conclusion that the participants ability to be situationally aware of threats was improved using conversational agents.

Hypothesis B: Performance of detecting threats will be improved using conversational agents (See hypotheses in Section 7.2.1).

Table 7.7 shows the mean detection efficiency of the **Verbal** (Ve) and **Aural** (Au) conversational agents compared to the baseline **Visual** (Vi) method. The detection efficiency when using a **Multi-modal** (Mu) approach was also reported.

The mean detection efficiency of the baseline **Visual** (Vi) (460.49) method was found to be lower than both the **Verbal** (Ve) (473.13) and **Aural** (Au) (539.52) agents, therefore **Hypothesis $\mathbf{H_5}$** was rejected.

The mean detection accuracy of the **Aural** agent (Au) (P = .869, R = .924, F_1 = .887) was lower than when using the baseline **Visual** (Vi) method (P = .973, R = .987, F_1 = .977), however, the results for the **Verbal** (Ve) agent (P = 1.0, R = 1.0, F_1 = 1.0) were found to be higher. As one agent was found to be more accurate **Hypothesis $\mathbf{H_6}$** was accepted.

The aim of this chapter was to explore the question "*Are conversational agents effective in making users situationally aware of threats in consumer IoT networks?*". The results showed that when participants used a **Multi-modal** (Mu) approach using a mix of conversational agent and the baseline visual method, accuracy (P = 1.0, R = 1.0, F_1 = 1.0) and efficiency (455.17) were both better than when using only the baseline visual method. It is, therefore, reasonable to conclude that participants performed better at detecting threats when using the conversational agents, demonstrating them to be an effective method of improving Cyber Situational Awareness.

7.5 Conclusions

This chapter, reported the results of a longitudinal study where the utility of conversational agents was assessed over an extended period lasting twenty-one days. The study was mapped to Mica Endsley's Situational Awareness model and was used to assess how participants perceive device activity, comprehend this in the context of their environment, and use the knowledge to determine if a threat exists. In addition, the usability of the agents was evaluated in terms of a users ability to achieve goals effectively, efficiently and with satisfaction. The study demonstrated that participants reported increased confidence in identifying threats when using the two agents. In addition, when using a multi-modal approach involving a combination of the conversational agents and the baseline visualisation tool, accuracy and efficiency were also improved. In the final chapter, a summary of the findings in this thesis are presented and the implications of the results are discussed. The limitations of the work are also discussed, before finally suggestions are provided about how the research could be extended and taken further.

Chapter 8

Conclusion

The aim of this thesis was to explore the use of conversational agents to improve Cyber Situational Awareness. This chapter begins by summarising the findings of this thesis and how these answer the research questions posed in Chapter 1. Second, it shows how the findings contribute to the body of knowledge in Cyber Situational Awareness. Third, it discusses the major strengths and limitations of each of the studies presented in this thesis, and then suggests avenues for potential future research.

8.1 Summary of Findings

In order to reflect on the contribution made in this thesis, the original research questions are revisited. Each question is considered in the context of the study conducted, and the contribution to knowledge each has made.

In Chapter 4 the study set out to answer the question "*Can current security methods detect the presence of threats within consumer IoT networks ?*". An existing open source IDS (*Snort*) which could be used freely within a consumer network was tested, and proved effective at detecting (*Mirai*) botnet traffic. However, the study also suggested that mutated versions of the malware could prove more difficult to detect, rendering existing signatures ineffective. A new threat detection model (BLSTM-RNN) was proposed which harnessed the power of deep learning for threat detection in consumer IoT networks. Once trained with previous attack data, the IDS model could accurately predict future threats facing consumer IoT network. At the time of undertaking this research, a lack of IoT botnet datasets was identified. An important output from this study is, therefore, the generation of a new *mirai* botnet dataset, providing a much needed resource for future research in this area. The dataset has already been made public and been used for comparative studies [14, 15].

In Chapter 5 the study changed the focus of detection from systems to users, and set out to answer the question "*Can users visually detect the presence of threats within consumer IoT networks?*". Understanding how users perceive risk, is an important consideration when attempting to evaluate and promote better situational awareness of risks relating to security and privacy. A cross-sectional study of 158 participants was undertaken to analyse the features users require from IoT devices and the importance placed upon security and privacy. The study also evaluated each users ability to identify if a smart device had been infected, and was being used to perform attacks on the Internet. Previous work [189, 190, 191] had suggested that demographic characteristics may have an effect on users awareness of threats. However, although this appeared to hold some truth, it was not evident whether a clear association existed between a users age or technical knowledge and their ability to detect threats. The results of the study demonstrated that users valued security and privacy but found identifying threats difficult. In addition, it found that a lack of information about network traffic can result in little or no awareness of security issues; however, if users were presented with data, awareness could be improved. It also clearly showed that the presentation of data is vitally important, otherwise the presence of the additional data, can have little impact. The research from the study contributes to the developing knowledge relating to risk perception and awareness. The contribution has significance since it tested assertions made in previous research, providing further clarity about their application to security and privacy within consumer IoT environments.

In Chapter 6 the study built upon the findings of the previous chapter to explore the viability of using conversational agents for improving situational awareness. It set out to answer the question "*Are conversational agents a viable method for making users aware of threats in consumer IoT networks?*" As shown in the literature review conversational agents have experienced a significant rise in popularity, and have been widely adopted by a range of companies, producing Microsoft's *Cortana*, Apple's *Siri*, Google's *Assistant* and Amazon's *Alexa*. A cross-sectional study of 72 participants was undertaken to assess the effectiveness of conversational agents for improving situational awareness. The study used a Pre-Study / Post-Study design, where participants indicated their agreement with confidence statements (mapped to Mica Endsley's SA model [1]) relating to their awareness and ability to monitor smart device and network activity. The statements were completed before and after using the agents to perform a series of tasks, allowing variance in Pre-Study/Post-Study scores to be observed. The results demonstrated that situational awareness was improved when using the two agents, since participants reported an increase in confidence in their ability to identify threats. In Chapter 7 the results of the previous cross-sectional study served as the basis for the final longitudinal study in this thesis. This study evaluated the utility of agents and

set out to answer the question "*Are conversational agents effective in making users situationally aware of threats in consumer IoT networks ?*". Sixteen participants took part in a twenty-one day study, which again was mapped to Mica Endsley's Situational Awareness model, and was used to assess how participants perceived device activity, comprehended this in the context of their environment, and used the knowledge to determine if a threat exists. In addition, the usability of the agents in terms of a users ability to achieve goals effectively, efficiently and with satisfaction, was evaluated. The results demonstrated that participants reported increased confidence in identifying threats when using the two agents. Importantly, the agents proved most effective when used as part of a multi-modal approach involving a combination of learning modalities (Aural, Verbal and Visual). The results also demonstrated that the use of the agents improved the accuracy and efficiency of detecting threats. The two studies presented in Chapter 6 and 7 provide a novel contribution to the developing body of knowledge, since collectively they build upon previous research such as [3, 189], which focused on improving Cyber Situational Awareness using a visual modality. These studies presented the use of aural and verbal modalities, demonstrating them also to be effective at improving Cyber Situational Awareness.

The ultimate objective of this thesis was to investigate if conversational agents could be used as a mechanism to improve Cyber Situational Awareness. In doing so, the research set out to answer the question "*Can Situational Awareness of threats in the Internet of Things be improved using Conversational Agents ?*". The results of the four studies clearly demonstrated that conversational agents can improve situational awareness of threats in the IoT. In doing so, this thesis has contributed to the body of knowledge by providing empirical evidence to add to the gaps in literature identified in Chapter 2. Namely, the use of deep learning (BSTM-RNN with word embedding) to extract semantic meaning from packets, and perform deep packet inspection. Also, providing the first study to investigate the effectiveness of conversational agents for threat detection and network monitoring, specifically the use of multi-modal agents to aid situational awareness of threats in consumer IoT networks.

8.2 Implications of Results

A key implication for research practice arising from this thesis surrounds the use of multi-modal approaches to improve awareness of security and privacy issues. The work has shown that the way information is presented has a major impact on the cognitive ability of users to understand events in their surroundings, which directly affects their ability to be situationally aware. The work has also demonstrated that Endsley's definition *"the perception of elements in the environment within a volume of time and*

space, the comprehension of their meaning, and the projection of their status in the near future" [1], still holds true, and can be successfully applied to improve situational awareness in a cyber context.

Additionally, and of importance when considering Smart Homes and the IoT, the work has shown the promise of using conversational agents to improve daily life. Ever since Alan Turing posed his question "*Can machines think?*" [125], the race has been on for developers to create the ultimate conversational experience. Competitions like the *Loebner Prize* and Amazon's *Alexa Prize* have encouraged developers to push the boundaries of artificial intelligence, the later in particular focusing on agents assuming the role of an assistant, conversing through fluent and enjoyable interactions, to help users in their daily lives. In this new era of voice computing, the growing popularity of conversational agents is clear, where personified AI is being widely adopted to control many aspects of home life, such as heating systems and smart appliances. This work has demonstrated how the benefits of using agents can be extended to address the growing issue of security and privacy within the home. The multi-modal approach of integrating conversational agents with existing technologies has implications for future research since it will likely lead to further questions not only in the area of Cyber Situational Awareness, but more broadly for questions relating to the IoT.

Finally, the work demonstrated how deep learning can be applied successfully to IoT security. In particular, to address the growing issue of malware and botnets targeting consumer IoT devices. The work has implications for future research in this area since it demonstrated that previous methods of deep learning are not restricted to technologies found within larger networks, but can also be applied to smart homes. Successful detection was achieved at the packet level, which could lead to further research using the methods in this thesis, and applying deep learning methods to networks unable to collect network traffic flows.

8.3 Limitations of Work

Despite promising results in all four studies, a number of limitations must be acknowledged relating to the research presented in this thesis.

In Chapter 4, the deep learning method of detecting IoT malware was only tested using a single type of IoT malware (*Mirai*). While it is acknowledged that testing included a range of DDoS attacks, they were all generated from the same malware. Using a different malware type, perhaps a phishing attack, may have returned different results. In addition, although previous research has used LSTM models to detect malware [97, 98, 99, 100], the studies all utilised flow traffic, not individual packet inspection.

Since the detection model presented in this thesis used a BLSTM-RNN with word embedding to extract semantic meaning from individual packets, comparisons with previous studies was not possible. Finally, at present the detection method utilises an offline detection method, meaning the full Conversational Cyber Situational Awareness Framework presented in Figure 7.3 has not yet been achieved.

For the study in Chapter 5 several limitations have also been identified. Firstly, the study is limited by the use of self-reported data. Since participants undertook the survey without interference from the researcher. Furthermore, socially desirable knowledge, skills and attitudes towards IoT security and privacy may have been provided and socially undesirable equivalents under reported. For example, participants may have ranked security and privacy highly because it is generally accepted as important, rather than because they personally believed this to be true. As a result, bias could have been introduced. In addition, the initial use of convenience sampling may have contributed to an over representation of student respondents (54%) in the total sample population. Since many were also studying a computing related course, this may also have contributed to the largest samples of technical knowledge levels being Intermediate (44%) and Advanced (36%). Finally, only one type of malware was again investigated. The use of other malware types may return different results, and provide a basis for further research in this area.

For the study in Chapters 6 and 7 similar limitations have been identified. Again, the use of convenience sampling in the cross-sectional study may have contributed to an over representation of student respondents (54%) in the total sample population. Since the study was held on the University campus, this could have resulted in the largest sample of participants being identified as aged 18-24 (47%)and with Advanced technical knowledge (44%). In addition, in the longitudinal study, while careful consideration was given to mitigating the Hawthorne Effect (See Section 7.2.2) the use of naturalistic observation may have introduced risk into the study, since the efficiency results could be less accurate if a participant was distracted while completing a task. This was partially mitigated by including pauses between tasks to allow for such distractions, and only recording time duration once a participant had started to undertake a task. However, the possibility of a distraction was still present.

8.4 Future Work

There are a number of possible avenues to further the work presented in this thesis. Firstly, some of the limitations of this research could be addressed. For instance, the novel approach of using a BLSTM-RNN with word embedding to extract semantic meaning from individual packets, has proven to be effective method of detecting IoT

malware. However, one avenue of further research would be to reproduce the model and test its ability to detect different types of malware. In addition, the model currently performs offline detection of malware, however this could be taken further to establish a real-time online detection mechanism. Finally, at the time of undertaking this research there was a lack of IoT malware datasets. Since then, new datasets may have been created, which could also be used to test the performance of future detection models based on the approach taken in this thesis.

Second, due to constraints of time and resources the longitudinal study presented in this thesis was limited to twenty-one days. Another avenue of further research would be to increase the number of participants and duration of the study. In addition, the verbal agent could be deployed to a range of different platforms, and the difference in their integration and performance tested. This thesis presents a novel study investigating the use of conversational agents to improve Cyber Situational Awareness. The research could be taken further by extending the agents to include more functionality and provide the user with more varied types of information. This could lead to further improvements in situational awareness, which could be measured and compared with the results in this research. Finally, the research highlighted the effectiveness of using a multi-modal approach to improve Cyber Situational Awareness. This aspect of the research could be extended to consider different combinations of modalities and explore if a clear association exists between a users preferred learning style, and the accuracy, effectiveness and satisfaction of using different conversational agents.

8.5 Final Remarks

In this thesis, the use of conversational agents was explored in the context of the IoT. In doing so, it provided the first analysis of their ability to improve Cyber Situational Awareness. The research demonstrated how deep learning could be used to detect IoT botnet activity in consumer IoT networks. Existing research has focused on detection in network flows [97, 98, 99], however, the method used in this research was the first to use a BLSTM-RNN with word embedding to extract semantic meaning from packets, and perform deep packet inspection to detect IoT malware. This was important since it is unlikely SOHO routers and consumer networks would be able to generate network flow traffic. The study resulted in a labelled dataset which has been made public and has already been used for comparative studies [14]. In addition, a cross-sectional study evaluated users awareness and perception of threats within the IoT, demonstrating that although users value security and privacy, they found it difficult to identify threats and infected devices. It also demonstrated that although a lack of network communication can result in little or no awareness of security issues; if users are presented with data,

their awareness could be improved. Finally, novel cross-sectional and longitudinal studies evaluated the use of conversational agents and demonstrated that agents could be used as an effective method to improve Cyber Situational Awareness. In particular, the study found this to be true when following Endsley's guideline of using a multi-modal approach [116], combing aural, verbal and visual modalities.

Bibliography

[1] Endsley M. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32-64. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1995 03;37:32–64.

[2] Onwubiko C. Understanding Cyber Situational Awareness. International Journal on Cyber Situational Awareness. 2016;1(1):11–30.

[3] Legg PA. Enhancing cyber situation awareness for Non-Expert Users using visual analytics. In: 2016 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (CyberSA); 2016. p. 1–8.

[4] McDermott CD, Petrovski AV. Investigation Of Computational Intelligence Techniques For Intrusion Detection In Wireless Sensor Networks. International Journal of Computer Networks and Communications (IJCNC). 2017;9(4):45–56.

[5] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Computer Networks. 2010;54(15):2787 – 2805.

[6] Ziegeldorf J, Morchon O, Wehrle K. Privacy in the Internet of Things: Threats and Challenges. Security and Communication Networks. 2014 12;7.

[7] McDermott CD, Isaacs JP, Petrovski AV. Evaluating Awareness and Perception of Botnet Activity within Consumer Internet-of-Things (IoT) Networks. Informatics. 2019;6(1).

[8] Henze M, Hermerschmidt L, Kerpen D, Häußling R, Rumpe B, Wehrle K. A comprehensive approach to privacy in the cloud-based Internet of Things. Future Generation Computer Systems. 2015 09;.

[9] Neisse R, Baldini G, Steri G, Ahmad A, Fourneret E, Legeard B. Improving Internet of Things device certification with policy-based management. Global Internet of Things Summit (GLoTS). 2017 06;p. 1–6.

[10] Sinanović H, Mrdovic S. Analysis of Mirai malicious software. In: 2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM); 2017. p. 1–5.

[11] McDermott CD, Haynes W, Petrovski AV. Threat Detection and Analysis in the Internet of Things using Deep Packet Inspection. International Journal on Cyber Situational Awareness. 2018;3(1):61–83.

[12] Bottom W, Gilovich T, Griffin D, Kahneman D. Heuristics and Biases: The Psychology of Intuitive Judgment. The Academy of Management Review. 2004 10;29:695.

[13] Newport C. Digital Minimalism On living Better with Less Technology. Penguin; 2019.

[14] Hwang RH, Peng MC, Nguyen VL, Chang YL. An LSTM-Based Deep Learning Approach for Classifying Malicious Traffic at the Packet Level. Applied Sciences. 2019 08;9:3414.

[15] Hwang RH, Peng MC, Huang CW, Lin PC, Nguyen VL. An Unsupervised Deep Learning Model for Early Network Traffic Anomaly Detection. IEEE Access. 2020 02;8:1–1.

[16] Verisign. Verisign Distributed Denial of Service Trends Report. Computer Networks. 2017;4.

[17] McDermott CD, Majdani F, Petrovski AV. Botnet Detection in the Internet of Things using Deep Learning Approaches. In: 2018 International Joint Conference on Neural Networks (IJCNN); 2018. p. 1–8.

[18] Jerkins JA. Motivating a market or regulatory solution to IoT insecurity with the Mirai botnet code. In: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC); 2017. p. 1–5.

[19] Ghiglieri M, Volkamer M, Renaud K. Exploring Consumers' Attitudes of Smart TV Related Privacy Risks. In: Tryfonas T, editor. Human Aspects of Information Security, Privacy and Trust. Cham: Springer International Publishing; 2017. p. 656–674.

[20] McDermott CD, Jeannelle B, Isaacs JP. Towards a Conversational Agent for Threat Detection in the Internet of Things. In: 2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA); 2019. p. 1–8.

[21] Mosenia A, Jha NK. A Comprehensive Study of Security of Internet-of-Things. IEEE Transactions on Emerging Topics in Computing. 2017 Oct;5(4):586–602.

[22] Kevin Ashton A. That Internet of things. RFiD Journal. 2009;.

[23] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems. 2013;29(7):1645 – 1660.

[24] Varaiya P. Smart cars on smart roads: problems of control. IEEE Transactions on Automatic Control. 1993;38(2):195–207.

[25] Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of Things for Smart Cities. IEEE Internet of Things Journal. 2014;1(1):22–32.

[26] Dorling K, Heinrichs J, Messier GG, Magierowski S. Vehicle Routing Problems for Drone Delivery. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017;47(1):70–85.

[27] Ackerman E, Strickland E. Medical delivery drones take flight in east africa. IEEE Spectrum. 2018;55(1):34–35.

[28] Baig MM, Gholamhooseini H. Smart Health Monitoring Systems: An Overview of Design and Modeling. Journal of Medical Systems. 2013;37.

[29] Cardile F, Iannizzotto G, Rosa FL. A vision-based system for elderly patients monitoring. In: 3rd International Conference on Human System Interaction; 2010. p. 195–202.

[30] Lin C, Lin P, Lu P, Hsieh G, Lee W, Lee R. A Healthcare Integration System for Disease Assessment and Safety Monitoring of Dementia Patients. IEEE Transactions on Information Technology in Biomedicine. 2008;12(5):579–586.

[31] Cheng HT, Zhuang W. Bluetooth-enabled in-home patient monitoring system: Early detection of Alzheimer's disease. IEEE Wireless Communications. 2010;17(1):74–79.

[32] Pandian P, Mohanavelu K, Safeer KP, Kotresh TM, Shakunthala DT, Gopal P, et al. Smart Vest: Wearable multi-parameter remote physiological monitoring system. Medical engineering physics. 2008 06;30:466–77.

[33] Scott JE, Scott CH. Drone Delivery Models for Healthcare. In: 2017 50th International Conference on System Sciences; 2017. p. 3297–3304.

[34] Choi-Fitzpatrick A, Chavarria D, Cychosz E, Dingens J, Duffey M, Koebel K, et al.. Up in the Air: A Global Estimate of Non-Military Drone Use: 2009-2015; 2016.

[35] Harrison C, Eckman B, Hamilton R, Hartswick P, Kalagnanam J, Paraszczak J, et al. Foundations for Smarter Cities. IBM Journal of Research and Development. 2010;54(4):1–16.

[36] Alavi AH, Jiao P, Buttlar WG, Lajnef N. Internet of Things-enabled smart cities: State-of-the-art and future trends. Measurement. 2018;129:589 – 606.

[37] Hui TKL, Sherratt RS, Sánchez DD. Major requirements for building Smart Homes in Smart Cities based on Internet of Things technologies. Future Generation Computer Systems. 2017;76:358 – 369.

[38] Nuortio T, Kytöjoki J, Niska H, Bräysy O. Improved route planning and scheduling of waste collection and transport. Expert Systems with Applications. 2006;30(2):223 – 232.

[39] Ahlgren B, Hidell M, Ngai EC. Internet of Things for Smart Cities: Interoperability and Open Data. IEEE Internet Computing. 2016;20(6):52–56.

[40] Atif Y, Ding J, Jeusfeld MA. Internet of Things Approach to Cloud-based Smart Car Parking. Procedia Computer Science. 2016;98:193 – 198.

[41] Stojkoska] BLR, Trivodaliev KV. A review of Internet of Things for smart home: Challenges and solutions. Journal of Cleaner Production. 2017;140:1454 – 1464.

[42] Mittal Y, Toshniwal P, Sharma S, Singhal D, Gupta R, Mittal VK. A voice-controlled multi-functional Smart Home Automation System. In: 2015 Annual IEEE India Conference (INDICON); 2015. p. 1–6.

[43] Lloret J, Macías E, Suárez A, Lacuesta R. Ubiquitous Monitoring of Electrical Household Appliances. Sensors. 2012;12.

[44] Jung M, Reinisch C, Kastner W. Integrating Building Automation Systems and IPv6 in the Internet of Things. In: 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing; 2012. p. 683–688.

[45] Harrison M, Vogt H, Kalaboukas K, Tomasella M, Wouters K, Gusmeroli S, et al. Vision and challenges for realising the Internet of Things. Cluster of European Research Projects on the Internet of Things - CERO IoT. 2010;.

[46] Aksu H, Babun L, Conti M, Tolomei G, Uluagac AS. Advertising in the IoT Era: Vision and Challenges. IEEE Communications Magazine. 2018;56(11):138–144.

[47] Aazam M, St-Hilaire M, Lung CH, Lambadaris I, Huh EN. In: Rahmani AM, Liljeberg P, Preden JS, Jantsch A, editors. IoT Resource Estimation Challenges and Modeling in Fog. Cham: Springer International Publishing; 2018. p. 17–31.

[48] Notra S, Siddiqi M, Habibi Gharakheili H, Sivaraman V, Boreli R. An experimental study of security and privacy risks with emerging household appliances. In: 2014 IEEE Conference on Communications and Network Security; 2014. p. 79–84.

[49] Schiefer M. Smart Home Definition and Security Threats. In: 2015 Ninth International Conference on IT Security Incident Management IT Forensics; 2015. p. 114–118.

[50] Moganedi S, Mtsweni J. Beyond the convenience of the internet of things: Security and privacy concerns. In: 2017 IST-Africa Week Conference (IST-Africa); 2017. p. 1–10.

[51] Sicari S, Rizzardi A, Grieco LA, Coen-Porisini A. Security, privacy and trust in Internet of Things: The road ahead. Computer Networks. 2015;76:146 – 164.

[52] Angrishi K. Turning Internet of Things(IoT) into Internet of Vulnerabilities (IoV) : IoT Botnets; 2017. ArXiv:1702.03681.

[53] De Donno M, Dragoni N, Giaretta A, Spognardi A. Analysis of DDoS-capable IoT malwares. In: 2017 Federated Conference on Computer Science and Information Systems (FedCSIS); 2017. p. 807–816.

[54] De Donno M, Dragoni N, Giaretta A, Spognardi A. DDoS-Capable IoT Malwares: Comparative Analysis and Mirai Investigation. Security and Communication Networks. 2018 02;2018:1–30.

[55] Keathley K. Security with IoT. In: Hammons RL, Kovac RJ, editors. Fundamentals of Internet of Things for Non-Engineers. CRC Press; 2019. .

[56] Marta Janus. Heads of the Hydra. Malware for Network Devices; 2011. https://securelist.com/heads-of-the-hydra-malware-for-network-devices/36396/, [Accessed on 2019-06-10].

[57] Vlajic N, Zhou D. IoT as a Land of Opportunity for DDoS Hackers. Computer. 2018;51(7):26–34.

[58] Mehic M, Slachta J, Voznak M. Whispering through DDoS attack. Perspectives in Science. 2016;7:95 – 100.

[59] Bohio MJ. Analyzing a Backdoor/Bot for MIPS Platform. SANS Institute; 2015. SECR/001/EN2001.

[60] Vignau B, Khoury R, Hallé S. 10 Years of IoT Malware: A Feature-Based Taxonomy. In: 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C); 2019. p. 458–465.

[61] Alomari E, Manickam S, Gupta BB, Karuppayah S, Alfaris R. Botnet-based Distributed Denial of Service (DDoS) Attacks on Web Servers: Classification and Art. International Journal of Computing Applications. 2012;47(7):24 – 32.

[62] Marzano A, Alexander D, Fonseca O, Fazzion E, Hoepers C, Steding-Jessen K, et al. The Evolution of Bashlite and Mirai IoT Botnets. In: 2018 IEEE Symposium on Computers and Communications (ISCC); 2018. p. 00813–00818.

[63] Angrishi K. Turning Internet of Things(IoT) into Internet of Vulnerabilities (IoV) : IoT Botnets. 2017 02;.

[64] Tom Spring. LizardStresser IoT Botnet Part of 400Gbps DDoS Attacks; 2016. https://threatpost.com/lizardstresser-iot-botnet-part-of-400gbps-ddos-attacks/119006/, [Accessed on 2019-06-12].

[65] Kambourakis G, Kolias C, Stavrou A. The Mirai botnet and the IoT Zombie Armies. In: 2017 IEEE Military Communications Conference (MILCOM); 2017. p. 267–272.

[66] Brian Krebs. KrebsOnSecurity Hit With Record DDoS.; 2016. https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/ , [Accessed on 2018-01-10].

[67] Hummel R, Hildebrand C, Modi H, Sockrider G, Dobbins R, Bjarnason S, et al. Netscout Threat Intelligence Report. Netscout; 2020. SECR/001/EN2001.

[68] Threat Intelligence Team. Seven new Mirai variants and the aspiring cybercriminal behind them; 2018. https://blog.avast.com/hacker-creates-seven-new-variants-of-the-mirai-botnet , [Accessed on 2019-03-21].

[69] L Cooley. The Evolution Of Mirai Botnet Source Code Presents Increased Risk Of Large-Scale DDoS Attacks; 2018. http://www.mondaq.com/unitedstates/x/732962/, [Accessed on 2018-11-28].

[70] Pierluigi Paganini. Mirai botnet evolution since its source code is available online; 2018. https://resources.infosecinstitute.com/mirai-botnet-evolution-since-its-source-code-is-available-online/ , [Accessed on 2018-11-28].

[71] Ruchna Nigam. New Mirai Variant Targets Enterprise Wireless Presentation Display Systems; 2019. https://unit42.paloaltonetworks.com/new-mirai-variant-targets-enterprise-wireless-presentation-display-systems/, [Accessed on 2019-11-12].

[72] Pascal Geenens. JenX – Los Calvos de San Calvicie; 2018. https://blog.radware.com/security/2018/02/jenx-los-calvos-de-san-calvicie/, [Accessed on 2019-11-12].

[73] Kolias C, Kambourakis G, Stavrou A, Voas J. DDoS in the IoT: Mirai and Other Botnets. Computer. 2017;50(7):80–84.

[74] Koli MS, Chavan MK. An advanced method for detection of botnet traffic using intrusion detection system. In: 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT); 2017. p. 481–485.

[75] Anderson JP. Computer Security Threat Monitoring and Surveillance; 1980.

[76] Denning DE. An Intrusion-Detection Model. IEEE Transactions on Software Engineering. 1987;SE-13(2):222–232.

[77] Lunt TF, Jagannathan R. A prototype real-time intrusion-detection expert system. In: Proceedings. 1988 IEEE Symposium on Security and Privacy; 1988. p. 59–66.

[78] Heberlein LT, Dias GV, Levitt KN, Mukherjee B, Wood J, Wolber D. A network security monitor. In: Proceedings. 1990 IEEE Computer Society Symposium on Research in Security and Privacy; 1990. p. 296–304.

[79] Mukherjee B, Heberlein LT, Levitt KN. Network intrusion detection. IEEE Network. 1994;8(3):26–41.

[80] Lippmann RP, Fried DJ, Graf I, Haines JW, Kendall KR, McClung D, et al. Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation. In: Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00. vol. 2; 2000. p. 12–26 vol.2.

[81] Bass T. Intrusion Detection Systems and Multisensor Data Fusion. Communications of The ACM. 2000 Apr;43(4):99–105.

[82] Zarpelão BB, Miani RS, Kawakani CT, [de Alvarenga] SC. A survey of intrusion detection in Internet of Things. Journal of Network and Computer Applications. 2017;84:25 – 37.

[83] Patcha A, Park JM. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks. 2007;51(12):3448 – 3470.

[84] Goebel J, Holz T. Rishi: Identify Bot Contaminated Hosts by IRC Nickname Evaluation. In: Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets. HotBots'07. USA: USENIX Association; 2007. p. 8.

[85] Gopal TS, Meerolla M, Jyostna G, Reddy Lakshmi Eswari P, Magesh E. Mitigating Mirai Malware Spreading in IoT Environment. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 2018. p. 2226–2230.

[86] Habibi J, Midi D, Mudgerikar A, Bertino E. Heimdall: Mitigating the Internet of Insecure Things. IEEE Internet of Things Journal. 2017;4(4):968–978.

[87] Gu G, Porras P, Yegneswaran V, Fong M, Lee W. BotHunter: Detecting Malware Infection through IDS-Driven Dialog Correlation. In: Proceedings of 16th USENIX Security Symposium. SS'07. USA: USENIX Association; 2007. .

[88] Hadi HJ, Sajjad SM, un Nisa K. BoDMitM: Botnet Detection and Mitigation System for Home Router Base on MUD. In: 2019 International Conference on Frontiers of Information Technology (FIT); 2019. p. 139–1394.

[89] Kumar A, Lim TJ. Early Detection of Mirai-Like IoT Bots in Large-Scale Networks through Sub-sampled Packet Traffic Analysis. In: Arai K, Bhatia R, editors. Advances in Information and Communication. Cham: Springer International Publishing; 2020. p. 847–867.

[90] Ben Said N, Biondi F, Bontchev V, Decourbe O, Given-Wilson T, Legay A, et al. Detection of Mirai by Syntactic and Behavioral Analysis. In: 2018 IEEE 29th International Symposium on Software Reliability Engineering (ISSRE); 2018. p. 224–235.

[91] Zhao D, Traore I, Sayed B, Lu W, Saad S, Ghorbani A, et al. Botnet detection based on traffic behavior analysis and flow intervals. Computers and Security. 2013;39:2–16.

[92] Meidan Y, Bohadana M, Mathov Y, Mirsky Y, Shabtai A, Breitenbacher D, et al. N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders. IEEE Pervasive Computing. 2018;17(3):12–22.

[93] Kumar CUO, Bhama PRKS. Detecting and confronting flash attacks from IoT botnets. The Journal of Supercomputing. 2019 10;.

[94] Wei Wang, Ming Zhu, Xuewen Zeng, Xiaozhou Ye, Yiqiang Sheng. Malware traffic classification using convolutional neural network for representation learning. In: 2017 International Conference on Information Networking (ICOIN); 2017. p. 712–717.

[95] Li C, Wang J, Ye X. Using a Recurrent Neural Network and Restricted Boltzmann Machines for Malicious Traffic Detection. NeuroQuantology. 2018 05;16:823–831.

[96] Yin C, Zhu Y, Fei J, He X. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. IEEE Access. 2017;5:21954–21961.

[97] Yuan X, Li C, Li X. DeepDefense: Identifying DDoS Attack via Deep Learning. In: 2017 IEEE International Conference on Smart Computing (SMARTCOMP); 2017. p. 1–8.

[98] Fu R, Zhang Z, Li L. Using LSTM and GRU neural network methods for traffic flow prediction. In: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC); 2016. p. 324–328.

[99] Radford BJ, Apolonio LM, Trias AJ, Simpson JA. Network Traffic Anomaly Detection Using Recurrent Neural Networks; 2018.

[100] Cui J, Long J, Min E, Mao Y. WEDL-NIDS: Improving Network Intrusion Detection Using Word Embedding-Based Deep Learning Method. In: Torra V, Narukawa Y, Aguiló I, González-Hidalgo M, editors. Modeling Decisions for Artificial Intelligence. Cham: Springer International Publishing; 2018. p. 283–295.

[101] Gu G, Zhang J, Lee W. BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic. In: Proceedings of the 15th Annual Network and Distributed System Security Symposium; 2008. .

[102] Gu G, Perdici R, Zhang J, Lee W. BotMiner: clustering analysis of network traffic for protocol and structure independent botnet detection. In: Proceedings of the 17th USENIX security symposium. USA; 2008. .

[103] Nobakht M, Sivaraman V, Boreli R. A Host-Based Intrusion Detection and Mitigation Framework for Smart Home IoT Using OpenFlow. In: 2016 11th International Conference on Availability, Reliability and Security (ARES); 2016. p. 147–156.

[104] Adat V, Gupta BB. A DDoS attack mitigation framework for internet of things. In: 2017 International Conference on Communication and Signal Processing (ICCSP); 2017. p. 2036–2041.

[105] Kasinathan P, Pastrone C, Spirito MA, Vinkovits M. Denial-of-Service detection in 6LoWPAN based Internet of Things. In: 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); 2013. p. 600–607.

[106] Kasinathan P, Costamagna G, Khaleel H, Pastrone C, Spirito MA. DEMO: An IDS Framework for Internet of Things Empowered by 6LoWPAN. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer Communications Security. CCS '13. New York, NY, USA: Association for Computing Machinery; 2013. p. 1337–1340.

[107] Ahmed Z, Danish SM, Qureshi HK, Lestas M. Protecting IoTs from Mirai Botnet Attacks Using Blockchains. In: 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD); 2019. p. 1–6.

[108] Javaid U, Siang AK, Aman MN, Sikdar B. Mitigating LoT Device Based DDoS Attacks Using Blockchain. In: Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems. CryBlock'18. New York, NY, USA: Association for Computing Machinery; 2018. p. 71–76.

[109] Rodrigues B, Bocek T, Lareida A, Hausheer D, Rafati S, Stiller B. A Blockchain-Based Architecture for Collaborative DDoS Mitigation with Smart Contracts. In: Tuncer D, Koch R, Badonnel R, Stiller B, editors. 11th IFIP International Conference on Autonomous Infrastructure, Management and Security (AIMS). vol. LNCS-10356 of Security of Networks and Services in an All-Connected World; 2017. p. 16–29.

[110] Bhardwaj K, Miranda JC, Gavrilovska A. Towards IoT-DDoS Prevention Using Edge Computing. In: USENIX Workshop on Hot Topics in Edge Computing (HotEdge 18). Boston, MA: USENIX Association; 2018. .

[111] Lee TH, Wen CH, Chang LH, Chiang HS, Hsieh MC. A Lightweight Intrusion Detection Scheme Based on Energy Consumption Analysis in 6LowPAN. In: Huang YM, Chao HC, Deng DJ, Park JJJH, editors. Advanced Technologies, Embedded and Multimedia for Human-centric Computing; 2014. p. 1205–1213.

[112] Shen W, Han G, Shu L, Rodrigues JJPC, Chilamkurti N. A New Energy Prediction Approach for Intrusion Detection in Cluster-Based Wireless Sensor Networks. In: Rodrigues JJPC, Zhou L, Chen M, Kailas A, editors. Green Communications and Networking. Berlin, Heidelberg; 2012. p. 1–12.

[113] Tadda GP, Salerno JS. Overview of Cyber Situation Awareness. In: Jajodia S, Liu P, Swarup V, Wang C, editors. Cyber Situational Awareness: Issues and Research. Boston, MA: Springer US; 2010. p. 15–35.

[114] Adams MJ, Tenney YJ, Pew RW. Situation Awareness and the Cognitive Management of Complex Systems. Human Factors. 1995;37(1):85–104.

[115] Alberts D, Garstka J, Hayes R, Signori D. Understanding Information Age Warfare. DoD Command and Control Research Program Pubication Series. 2001 08;p. 320.

[116] Stanton N, Chambers P, Piggott J. Situational awareness and safety. Safety Science. 2001 12;39:189–204.

[117] Endsley MR. Design and Evaluation for Situation Awareness Enhancement. Proceedings of the Human Factors Society Annual Meeting. 1988;32(2):97–101.

[118] Bedny G, Meister D. Theory of Activity and Situation Awareness. International Journal of Cognitive Ergonomics. 1999;3(1):63–72.

[119] Smith K, Hancock P. Situation Awareness Is Adaptive, Externally Directed Consciousness. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1995 03;37:137–148.

[120] McGuiness B, Foy JL. A subjective measure of SA: The crew awareness rating scale (cars); 2000. p. 286–291.

[121] Hall DL, Llinas J. An introduction to multisensor data fusion. Proceedings of the IEEE. 1997 Jan;85(1):6–23.

[122] Onwubiko C. Functional requirements of situational awareness in computer network security. In: 2009 IEEE International Conference on Intelligence and Security Informatics; 2009. p. 209–213.

[123] Abushawar B, Atwell E. ALICE chatbot: Trials and outputs. Computación y Sistemas. 2015 12;19.

[124] Furey E, Blue J. She Knows Too Much – Voice Command Devices and Privacy. In: 2018 29th Irish Signals and Systems Conference (ISSC); 2018. p. 1–6.

[125] TURING AM. Computing Machinery and Intelligence. Mind. 1950 10;LIX(236):433–460.

[126] Zemčík T. A Brief History of Chatbots. In: 2019 International Conference on Artificial Intelligence, Control and Automation Engineering (AICAE 2019); 2019. .

[127] Weizenbaum J. ELIZA: A Computer Program for the Study of Natural Language Communication between Man and Machine. Commun ACM. 1966 Jan;9(1):36–45.

[128] Colby KM. In: Wilks Y, editor. Human-Computer Conversation in A Cognitive Therapy Program. Boston, MA: Springer US; 1999. p. 9–19.

[129] Ferrucci D, Brown E, Chu-Carroll J, Fan J, Gondek D, Kalyanpur A, et al. Building Watson: An Overview of the DeepQA Project. AI Magazine. 2010 09;31:59–79.

[130] Abdul-Kader SA, Woods J. Survey on Chatbot Design Techniques in Speech Conversation Systems. International Journal of Advanced Computer Science and Applications;6(7).

[131] Vlahos J. Talk To Me. Amazon, Google, Apple and the race for Voice Controlled AI. R H Business Books; 2019.

[132] Hoy MB. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Medical Reference Services Quarterly. 2018;37(1):81–88.

[133] Gao Y, Pan Z, Wang H, Chen G. 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); 2018. p. 372–380.

[134] Sandbank T, Shmueli-Scheuer M, Herzig J, Konopnicki D, Richards J, Piorkowski D. Detecting Egregious Conversations between Customers and Virtual Agents. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, Louisiana: Association for Computational Linguistics; 2018. .

[135] Io HN, Lee CB. Chatbots and conversational agents: A bibliometric analysis. In: 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); 2017. p. 215–219.

[136] Paikari E, van der Hoek A. A Framework for Understanding Chatbots and Their Future. In: 2018 IEEE/ACM 11th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE); 2018. p. 13–16.

[137] Rajalakshmi A, Shahnasser H. Internet of Things using Node-Red and alexa. In: 2017 17th International Symposium on Communications and Information Technologies (ISCIT); 2017. p. 1–4.

[138] Yue CZ, Ping S. Voice activated smart home design and implementation. In: 2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST); 2017. p. 489–492.

[139] Solorio JA, Garcia-Bravo JM, Newell BA. Voice Activated Semi-Autonomous Vehicle Using Off the Shelf Home Automation Hardware. IEEE Internet of Things Journal. 2018 Dec;5(6):5046–5054.

[140] Van Cuong T, Tan TM. Design and Implementation of Chatbot Framework For Network Security Cameras. In: 2019 International Conference on System Science and Engineering (ICSSE); 2019. p. 324–328.

[141] Këpuska V, Bohouta G. 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); 2018. p. 99–103.

[142] Reis A, Paulino D, Paredes H, Barroso I, Monteiro MJ, Rodrigues V, et al. 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); 2018. p. 1–5.

[143] Kerry A, Ellis R, Bull S. Conversational Agents in E-Learning. In: Allen T, Ellis R, Petridis M, editors. Applications and Innovations in Intelligent Systems XVI. London: Springer London; 2009. p. 169–182.

[144] Shepherd LA, Archibald J. Security awareness and affective feedback: Categorical behaviour vs. reported behaviour. In: 2017 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA); 2017. p. 1–6.

[145] Shepherd LA, Archibald J, Ferguson RI. Assessing the Impact of Affective Feedback on End-User Security Awareness. In: Tryfonas T, editor. Human Aspects of Information Security, Privacy and Trust. Cham: Springer International Publishing; 2017. p. 143–159.

[146] Sacharin V, Sander D, Scherer KR. The perception of changing emotion expressions. Cognition and Emotion. 2012;26(7):1273–1300. PMID: 22550942.

[147] Onwubiko C. Understanding Cyber Situation Awareness. International Journal on Cyber Situational Awareness. 2016;1(1):11–30.

[148] Rosenthal R, Rosnow RL. Essentials of Behavioral Research: Methods and Data Analysis. Third edition ed. New York: McGraw-Hill.; 2008.

[149] Lazar J, Feng JH, Hochheiser H. Research Methods in Human-Computer Interaction. 2nd ed. Cambridge, MA: Morgan Kaufmann; 2017.

[150] Zukauskas P, Vveinhardt J, Andriukaitienė R. Philosophy and Paradigm of Scientific Research. In: Management Culture and Corporate Social Responsibility. Rijeka: IntechOpen; 2018. .

[151] Orlikowski W, Baroudi J. Studying Information Technology in Organizations: Research Approaches and Assumptions. Information Systems Research. 1991 03;2:1–28.

[152] Creswell JW. Research design: Qualitative, quantitative, and mixed methods approaches. Sage; 2009.

[153] Macefield R, Close C, Nu WW. How To Specify the Participant Group Size for Usability Studies: A Practitioner's Guide. Journal of Usability Studies. 2009;p. 34–45.

[154] Virzi RA. Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? Human Factors. 1992;34(4):457–468.

[155] Nielsen J, Landauer TK. A Mathematical Model of the Finding of Usability Problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery; 1993. p. 206–213.

[156] Granello DH, Wheaton JE. Online Data Collection: Strategies for Research. Journal of Counseling & Development. 2004;82(4):387–393.

[157] Oehlert GW. A First Course in Design and Analysis of Experiments; 2000. .

[158] Sears A, Feng J, Oseitutu K, Karat CM. Hands-Free, Speech-Based Navigation during Dictation: Difficulties, Consequences, and Solutions. Human Computer Interaction. 2003 Sep;18(3):229–257.

[159] Pesonen HL, Ekvall T, Fleischer G, Huppes G, Jahn C, Klos Z, et al. Framework for scenario development in LCA. The International Journal of Life Cycle Assessment. 2000 04;5:21–30.

[160] Babbie E. Survey Research Methods. 2nd ed. Belmont, CA: Wadsworth Publishing; 1990.

[161] Dillman D. Mail and Internet Surveys: The Tailored Design Method. vol. 2; 2000.

[162] McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. Journal of Clinical Epidemiology. 2014;67(3):267 – 277.

[163] Ahuja K, Gulshan D. Evaluation Metrics for Intrusion Detection Systems-A Study. International Journal of Computer Science and Mobile Applications. 2015 06;11.

[164] Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry. CRC Press; 1996.

[165] Tong Z, Weiss SM. The handbook of data mining. Lawrence Erlbaum Assoociates; 2003.

[166] Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction. 2008;24(6):574–594.

[167] Adamson KA, Prion S. Reliability: Measuring Internal Consistency Using Cronbach's . Clinical Simulation in Nursing. 2013;9(5):e179 – e180.

[168] Zaharias P, Poylymenakou A. Developing a usability evaluation method for e-learning applications: Beyond functional usability. International Journal of Human-Computer Interaction. 2009;25(1):75–98.

[169] Laake P, Fagerland MW. Chapter 11 - Statistical Inference. In: Laake P, Benestad HB, Olsen BR, editors. Research in Medical and Biological Sciences (Second Edition). second edition ed. Amsterdam: Academic Press; 2015. p. 379 – 430.

[170] Riffenburgh RH. Chapter 9 - Tests on Categorical Data. In: Riffenburgh RH, editor. Statistics in Medicine (Third Edition). third edition ed. San Diego: Academic Press; 2012. p. 175 – 202.

[171] Williams LL, Quave K. Chapter 10 - Tests of Proportions: Chi-Square, Likelihood Ratio, Fisher's Exact Test. In: Williams LL, Quave K, editors. Quantitative Anthropology. Academic Press; 2019. p. 123 – 141.

[172] Scheff SW. Chapter 8 - Nonparametric Statistics. In: Scheff SW, editor. Fundamental Statistical Principles for the Neurobiologist. Academic Press; 2016. p. 157 – 182.

[173] Hoffman JIE. Chapter 26 - Analysis of Variance. II. More Complex Forms. In: Hoffman JIE, editor. Basic Biostatistics for Medical and Biomedical Practitioners (Second Edition). second edition ed. Academic Press; 2019. p. 419 – 441.

[174] Brooks J, Bartys S, Turley E, King N. The Utility of Template Analysis in Qualitative Psychology Research. Qualitative Research in Psychology. 2015 04;12.

[175] Brooks J, King N. Doing Template Analysis: Evaluating an End of Life Care Service. Sage Research Methods Cases. 2014 January;.

[176] Azar B. Online experiments: Ethically fair or foul? Monitor on Psychology. 2000;4.

[177] Finn P, Jakobsson M. Designing ethical phishing experiments. IEEE Technology and Society Magazine. 2007 02;26:46 – 58.

[178] Elzen IV, Heugten JV. Techniques for Detecting Compromised IoT Devices. University of Amsterdam; 2017.

[179] Khraisat A, Gondal I, Vamplew P, Kamruzzaman J. Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity. 2019 12;2.

[180] Heenan R, Moradpoor N. A survey of Intrusion Detection System technologies. In: PGCS 2016: The First Post Graduate Cyber Security Symposium; 2016. .

[181] Albin E, Rowe N. A Realistic Experimental Comparison of the Suricata and Snort Intrusion-Detection Systems. In: 2012 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA); 2012. p. 122–127.

[182] Day D, Burns BM. A performance analysis of Snort and Suricata Network Intrusion Detection and Prevention Engines. In: ICDS 2011, The Fifth International Conference on Digital Society. IARIA; 2011. p. 187–192.

[183] Wang X, Kordas A, Hu L, Gaedke M, Smith D. Administrative evaluation of intrusion detection system. In: RIIT 2013 - Proceedings of the 2nd Annual Conference on Research in Information Technology; 2013. p. 47–52.

[184] Ray A, Rajeswar S, Chaudhury S. Text recognition using deep BLSTM networks. In: 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR); 2015. p. 1–6.

[185] P PW, Qian Y, Song F, He L, Zhao H. A Unified Tagging Solution: Bidirectional LSTM Recurrent Neural Network with Word Embedding. ArXiv e-prints. 2015 Nov;.

[186] HaddadPajouh H, Dehghantanha A, Khayami R, Choo KKR. A deep Recurrent Neural Network based approach for Internet of Things malware threat hunting. Future Generation Computer Systems. 2018;85:88 – 96.

[187] Hamed T, Dara R, Kremer SC. Chapter 6 - Intrusion Detection in Contemporary Environments. In: Vacca JR, editor. Computer and Information Security Handbook (Third Edition). third edition ed. Boston: Morgan Kaufmann; 2017. p. 109 – 130.

[188] Liao HJ, Lin] CHR, Lin YC, Tung KY. Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications. 2013;36(1):16 – 24.

[189] Legg PA. Visualizing the insider threat: challenges and tools for identifying malicious user activity. In: 2015 IEEE Symposium on Visualization for Cyber Security (VizSec); 2015. p. 1–7.

[190] Vergelis, Maria. Older and Wiser ? A Look at the threats faced by over 55 online; 2016. https://media.kasperskycontenthub.com/wp-content/uploads/sites/45/2018/03/08233606/Report_Over-55s_Online_ENG.pdf, [Accessed on 2018-03-10].

[191] Amran A, Zaaba ZF, Singh MM, Marashdih AW. Usable Security: Revealing End-Users Comprehensions on Security Warnings. Procedia Computer Science. 2017;124:624– 631. 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia.

[192] Yamauchi M, Ohsita Y, Murata M, Ueda K, Kato Y. Anomaly Detection for Smart Home Based on User Behavior. In: 2019 IEEE International Conference on Consumer Electronics (ICCE); 2019. p. 1–6.

[193] Miller R, Wright D. Detecting and Correcting Attrition Bias in Longitudinal Family Research. Journal of Marriage and the Family. 1995 11;57:921.

[194] Amico KR. Percent Total Attrition: A Poor Metric for Study Rigor in Hosted Intervention Designs. American journal of public health. 2009 08;99:1567–75.

[195] Dienlin T, Trepte S. Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. European Journal of Social Psychology. 2015;45:285 – 297.

[196] Barth S, De Jong M. The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review. Telematics and Informatics. 2017 04;.

[197] Parsons K, McCormac A, Pattison M, Butavicius M, Jerram C. An Analysis of Information Security Vulnerabilities at Three Australian Government Organisations. In: 2013 European Information Security Multi-Conference. Springer International Publishing; 2013. p. 34–44.

[198] Potzsch S. Privacy Awareness: A Means to Solve the Privacy Paradox? In: The Future of Identity in the Information Society. Springer Berlin Heidelberg; 2009. p. 226–236.

[199] Onwubiko C, Owens T. Review of Situational Awareness for Computer Network Defense. In: Onwubiko C, Owens T, editors. Situational Awareness in Computer Network Defense: Principles, Methods and Applications; 2012. p. 1–9.

[200] Shin J, Guo Q, Gierl MJ. Multiple-Choice Item Distractor Development Using Topic Modeling Approaches. Frontiers in Psychology. 2019;10:825.

[201] Wu M, Miller RC, Garfinkel SL. Do Security Toolbars Actually Prevent Phishing Attacks? In: Grinter R, Rodden T, Aoki P, editors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2006. p. 601–610.

[202] Egelman S, King J, Miller RC, Ragouzis N, Shehan E. Security User Studies: Methodologies and Best Practices. In: CHI '07 Extended Abstracts on Human Factors in Computing Systems; 2007. p. 2833–2836.

[203] Deke J, Chiang H. The WWC Attrition Standard: Sensitivity to Assumptions and Opportunities for Refining and Adapting to New Contexts. Evaluation Review. 2017;41(2):130–154.

[204] Koksal, Ilker. Who's the Amazon Alexa Target Market, Anyway ?; 2018. https://www.forbes.com/sites/ilkerkoksal/2018/10/10/whos-the-amazon-alexa-target-market-anyway/#680589e92eb5, [Accessed on 2019-01-15].

[205] Connelly LM. Pilot studies. Medsurg nursing : official journal of the Academy of Medical-Surgical Nurses. 2008 December;17(6):411—412.

[206] Hill R. What Sample Size is "Enough" in Internet Survey Research ? Interpersonal Computing and Technology: An Electronic Journal for the 21st Century. 2008 July;6:3–4.

[207] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.

[208] Landsberger HA. Hawthorne Revisited. Management and the worker: its critics, and developments in human relations in industry. Cornell University. 1958;.

[209] McCarney R, Warner J, Iliffe S, Haselen R, Griffin M, Fisher P. The Hawthorne Effect: A Randomised, Controlled Trial. BMC medical research methodology. 2007 02;7:30.

[210] Mestre LS. The value and process of usability studies. In: Mestre LS, editor. Designing Effective Library Tutorials. Chandos Learning and Teaching Series. Chandos Publishing; 2012. p. 223 – 246.

[211] Arachchilage NAG, Love S. A game design framework for avoiding phishing attacks. Computers in Human Behavior. 2013;29(3):706 – 714.

[212] Parsons K, McCormac A, Butavicius M, Pattinson M, Jerram C. Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q). Computers Security. 2014;42:165 – 176.

[213] Parsons K, Calic D, Pattinson M, Butavicius M, McCormac A, Zwaans T. The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. Computers Security. 2017;66:40 – 51.

Appendix A

Sand-boxed Environment (Chapter 4)

A summary of the procedure undertaken to setup the main components of the sand-boxed botnet environment used in Chapters 4-7 is presented.

C&C Server Configuration

Essential packages were installed using *apt-get install unzip gcc golang electric-fence screen –y*

Domains were created for report.McDPhD.org and cnc.McDPhD.org, and added to *table.c* and *main.go*.

MySQL was installed using *apt-get install mysql-server mysql-client –y* and a user created using *INSERT INTO users VALUES (NULL, 'miraiuser', 'miraipassword', 0, 0, 0, 0, -1, 1, 30, ”)*; Once configured *main.go* was edited to include the MySQL credentials.

Cross compilers for the required binary architectures (e.g. *arm*, *mips*) were installed and appropriate export paths added to */etc/profile* using *export PATH= $PATH: /etc/x-compile/mips/bin*. To allow information regarding C&C connections, compiler issues and flood status to be sent the C&C server *./build.sh debug telnet* was run. The required binary files for each architecture were created and stored in the *release* directory using *./build.sh release*

Scan Loader Server Configuration

Apache was installed using *apt-get install apache2 –y* and binary architecture files created earlier, were moved to the *loader/bins* directory. The Scan/Loader IP address was added to *main.c* and full permission granted using *chmod777**. The *loader* file was compiled and added to the loader directory using *./build.sh*

To reduce the number of IP ranges available for scanning and ensure the range used in the environment was allowed, excluded IP ranges were amended in *scanner.c* to reflect the topology.

The Scan/Loader IP address was added to *scanListen.go* and port *48101* specified as the default port to listen for brute force results. Within the *tools* directory the *scanListen* file was compiled using *go build scanListen.go* and moved to the *loader* directory.

The Sricam AP009 IP camera used in the lab setup did not include *wget*, therefore tftp was installed using *apt-get install tftpd tftp*.

A tftp configuration was created using *touch /etc/xinetd.d/tftp* and */tftpboot* specified as the directory where the architecture binary files will be copied to for delivering later delivering the payload.

DNS Server Configuration

The Mirai malware requires access to a DNS server to discover the C&C server's IP address. *Bind9* software was installed and used to create two required domains *report.McDPhD.org* and *cnc.McDPhD.org* in *named.conf.local*. These will be used by the bots to report IoT device information and communicate with the C&C server.

Appendix B

Conversational Agent Architecture (Chapters 6 and 7)

A summary of the procedure undertaken to setup the conversational agents used in Chapters 6-7 is presented. The architecture consisted of three components; a data pipeline (1) which uploaded classified IDS logs into the back end *DynamoDB* table (2) via an *S3 bucket* in Amazon Web Services, and front end architectures (3) built for both agents.

1. ETL Pipeline

The implementation of the ETL pipeline required three processes, as shown in Figure 6.2. 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 classified *IDS* logs, and invoked a process to upload newly added *JSON* files to an *S3* bucket on AWS. For the studies in Chapters 6-7, the dataset created in Chapter 4 was manually added to the directory and processed by the ETL pipeline.

First, to handle the backend functionality, an *IAM Role* was required. From the AWS Management Console, a new *IAM Role* was created and (*AmazonS3FullAccess*, *AmazonDynamoDBFullAccess*, and *AWSOpsWorksCloudWatchLogs*) permissions assigned. In the pipeline, a Lambda Function *(step 2)* was used to transfer items from the *S3* bucket into DynamoDB *(step 3)*. A table was created using the attributes found in the *JSON* dataset (see Source Code 6.1). DynamoDB is a schema-less database that only requires a table name and primary key. The table's 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.

2. Frontend Architectures

Aural (Au) Agent

The Alexa Developer Console[1] was used to create the frontend for the **Aural** aegnt. A new skill named *Threat Detector* was created, an invocation name was assigned, and used to invoke the Alexa Skill from the Echo device. Twelve custom intents were configured, and used to trigger specific event functionality and enable a user to query the *DynamoDB* table for information. Seven in-built intents were used as triggers to perform preconfigured functionality such as *repeat*, *stop* or *cancel* an intent. The twelve intents are 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.
6. ***networkStatusToday:*** Responds to a user query and informs if there has been any issues detected on the network.
7. ***listSrcDevToday:*** Responds to a user query and returns a list of all active source devices on the network today.
8. ***listSrcDevByDate:*** Responds to a user query and returns a list of all active source devices on a specified date.

[1] https://developer.amazon.com/en-GB/alexa

9. ***activityTotalBySrcDevLastThreeDays:*** Responds to a user query and returns details of how much activity a specified source device had on each of the last three days.

10. ***activityTotalLastThreeDays:*** Responds to a user query and returns a summary of how much activity has occurred on each of the last three days.

11. ***unusualActivityLastThreeDays:*** Responds to a user query and returns a summary of normal and unusual activity on each of the last three days.

12. ***mostActiveSrcDevLastThreeDays:*** Responds to a user query and returns a list of three most active source devices on each of the last three days.

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 represented variable information a user specified, 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 was 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.

Verbal (Ve) Agent

For consistency, frontend development for the **Verbal** agent followed a similar methodology as the **Aural** agent however, Dialogflow[1], a google platform for creating conversational agents was used. A new agent called *Threat Detector* was created and the same twelve custom intents were configured to enable a user to query the *DynamoDB* table for information. These were combined with in-built intents such as *Default Fallback Intent* and *Default Welcome Intent* and used to trigger specific event functionality.

For each custom intent the same *utterances* were created but amended to be text based queries rather than converted speech, in order to help the agent identify the user's intent. Again, utterances which contained words that represented variable information

[1] https://dialogflow.cloud.google.com/

a user specified, were assigned an *Action*. The action functioned the same as the *slot* for the **Aural** agent and specified a placeholder within a sentence to be replaced by the specific variable information requested by the user.

We again needed to set the endpoint as *AWS Lambda* so that user queries could be serviced by the *Lambda Function* in order to query the *DynamoDB* table. To link Google's Dialogflow with Amazons Web Services we created an *AWS API*, and set the fulfillment of each intent to use this webhook and send user queries to *AWS Lambda*.

Finally, we deployed the **Verbal** agent as a chatbot on the *Telegram* messaging platform.

3. Backend Architecture

The main components of the backend architecture are the *AWS Lambda* function and *DynamoDB* table (See Figure 6.3). 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. The DynamoDB created earlier was used to store all uploaded IDS logs.

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. The *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.

Appendix C

Conversational Agent Intents (Chapters 6 and 7)

Table C.1: Conversational Agent Intents

	Intent	Response
i1	**activitySummaryToday** **Example:** *give me a summary of today's activity*	Alright, Today there has been 66.19 percent normal activity and 33.81 percent unusual activity. Anything else I may help you with?
i2	**activitySummaryByDate** **Example:** *give me a summary of activity for 20 january 2019*	Alright, On 2019-01-20 there was 66.19 percent normal activity and 33.81 percent unusual activity. Any other activities you want know about?
i3	**activitySummarySrcDevAndDate** **Example:** *give me a summary of activity by Smart Camera on 20 january 2019*	Alright, Summary for 2019-01-20 For the device, smart camera there was 6.25 percent normal activity and 93.75 percent unusual activity. Anything else I may help you with?
i4	**firstUnusualActivityByDate** **Example:** *give me detail of first unusual activity on 20 january 2019*	Alright, First unusual activity on 2019-01-20 For which, I.D is 690 Date Time is 20/01/2019 09:53 Source Device is Smart Camera Destination Device is Smart Fridge Data Type is UDP Activity type is unusual. Anything else I may help you with?
i5	**activityDetailsByID** **Example:** *give me details of activity id 366*	I.D is 366 Date for that is 2019-01-18 time is 19:39 Source Device is Amazon Echo Destination Device is Smart Camera Data Type is ARP Activity type is normal. What else would you like to know?
i6 ±	**networkStatusToday** **Example:** *what is the status of the network today*	There has been issues detected on the network today
i7 ±	**listSrcDevToday** **Example:** *give me a list of all active source devices today*	Today Source Device Smart Doorbell , Tablet , Amazon Echo , Smart Camera , Smart Fridge , Smart Thermostat , Smart Bulb , and Laptop have been active on the network.
i8 ±	**listSrcDevByDate** **Example** *give me a list of all active source devices on 21/01/2019*	On 2019-01-21 source device Smart Camera was active on the network.
i9 ±	**activityTotalBySrcDevLastThreeDays** **Example:** *give me activity totals for Smart Camera for the last three days*	Alright, On 2019-01-19 source device Smart Camera had 2 normal activities and 30 unusual activities. On 2019-01-18 source device Smart Camera had 67 normal activities and 23 unusual activities. On 2019-01-17 source device Smart Camera had 26 normal activities and 90 unusual activities.
i10 ±	**activityTotalLastThreeDays** **Example:** *give me total network activity for the last three days*	Alright, On 2019-01-19 there were 2 normal activities and 30 unusual activities. Then 2019-01-18 had 330 normal activities and 23 unusual activities. And 2019-01-17 had 110 normal activities and 90 unusual activities.
i11 ⊕	**unusualActivityLastThreeDays** **Example:** *has there been any unusual activity in the last three days*	Yes, there has been 24.44 percent unusual activity in last three days. It was detected on 2019-01-19 , 2019-01-18 and 2019-01-17
i12 ⊕	**mostActiveSrcDevLastThreeDays** **Example:** *give me a list of the most active source devices for the last three days*	The most active source devices on 2019-01-19 were and Smart Camera . The most active source devices on 2019-01-18 were Smart Doorbell , Smart Camera , and Smart Fridge . The most active source devices on 2019-01-17 were Smart Camera , Smart Doorbell , and Smart Fridge.

± *Added in Chapter 6 Main Study*, ⊕ Added in *Chapter 7 Study 1-4*

Appendix D

Use-Cases and Scenarios (Chapters 6 and 7)

The use-cases used to evaluate the use of conversational agents for improving situational awareness are presented. Each use-case was designed to represent realistic descriptions of how a user might want to use the conversational agents for monitoring smart device and network activity. To avoid introducing bias a different set of scenarios were created for each study in Chapters 6 and 7 based on the nine use-cases described in Section 3.3.2.

Table D.1: Scenarios used in Main Study Chapter 6

		Scenario
uc1	**sc1**	You have returned from work and would like a summary of network activity to check if any unusual activity has taken place on your smart devices, while you were away.
uc2	**sc2**	You have returned from a weekend away and suspect one of your smart devices may have been compromised, causing unusual activity.
uc3	**sc3**	It was reported yesterday on the Internet, that a model of smart camera (which you own) was found to contain a vulnerability, resulting in it being used to perform attacks on the Internet.
uc4	**sc4**	If you found unusual activity for Use-Case 3 it would be useful for you to know when the attack first took place.
uc5	**sc5**	You have identified that your Smart Fridge has also been compromised, and would like to know where on the Internet attacks are being sent.
uc6 ±	**sc6**	Concerned that other smart devices may have been compromised, you would like to know if any other smart devices have been active on the network for the last two days, without your knowledge.
uc7 ±	**sc7**	It has been a week since your Smart Camera was compromised and subsequently fixed. You would now like to quickly check if total daily network activity for the last two days has been normal, or unusually high.

± *not used in pilot study*

Table D.2: Scenarios used in Study 1 Chapter 7

		Day 1	Day 2	Day 3
uc1	**sc8**	Your network and smart devices appear to be functioning normaly, however you would like a quick summary of network activity to check if any unusual activity has taken place on your smart devices, while you were away.		
uc2	**sc11**		You have returned from a weekend away and suspect one of your smart devices may have been compromised, causing unusual activity. You would like to check if any unusual activity occured on the network yesterday.	
uc3	**sc12**		Having established that there has been unusual activity on the network today, you would like to check the usage of the smart camera yesterday.	
uc4	**usc10**	You suspect more than one smart device on your network has been compromised, and would like to know which device had the first unusual activity today.		
uc5	**sc13**		You have identified which smart devices have been compromised, and would like to know more information about a specific activity on the network	
uc6	**sc9**	Having established that there has been unusual activity on the network today, you would like to know which smart devices have been active on the network.		
uc7	**sc15**			You believe the rest of the smart devices on your network are now functioning normally. To confirm this you would like to know the total network activity for the previous two days.
uc8	**sc14**			Your previously compromised Smart Camera has been fixed. You would now like to quickly check if the activity level for the Smart Camera has been normal, or unusually high, for the previous two days.
uc9	**sc16**			You would like to continue monitoring specific devices for a few days, to establish what their normal network activity looks like. First you would like to know which devices have been most active on the network.

Table D.3: Scenarios used in Study 2 Chapter 7

		Day 1	Day 2	Day 3
uc1	**sc17**	You have returned from work and would like a summary of network activity to check if any unusual activity has taken place on your smart devices, while you were away.		
uc2	**sc20**		You have returned from a business trip and would like to check your smart devices for any unusual activity while you were away.	
uc3	**sc21**		Having established that there has been unusual activity on the network today, you would like to check the usage of the Smart Camera.	
uc4	**sc19**	Having established that unusual activity has occurred on the network, it would be useful to know more information about the first unusual activity which took place.		
uc5	**sc22**		You are confident there has been unusual activity on the network and would like to investigate further. You have identified when unusual activity may have occurred, and would like to check specific activity on the network.	
uc6	**sc18**	Concerned that other smart devices may have been compromised, you would like to know if any other smart devices have been active on the network today, without your knowledge.		
uc7	**sc24**			You suspect that activity on your network has increased, possibly due to another smart device being compromised. You would therefore like to quickly check if activity has increased over the past few days.
uc8	**sc23**			Your network appears to have been functioning normally for several days, having previously had issues with some smart devices. You would like to be pro-active and establish how much your smart camera normally uses your network, so you can spot future issues more quickly.
uc9	**sc25**			You recently completed an online shopping order with Tesco, and used your Smart Fridge to identify which food items you required. However, you completed the whole transaction from a public wifi hotspot, and since returning home you have noticed unusual items in the shopping list. You suspect you may have put the Smart Fridge at risk and would like to compare its activity with other smart devices for the previous two days.

Table D.4: Scenarios used in Study 3 Chapter 7

		Day 1	Day 2	Day 3
uc1	**sc26**	You have recently added two new smart devices to your home network. All devices appear to be functioning normally, however you would like to quickly check and ensure the network is functioning normally.		
uc2	**sc29**		It has been several days since you last checked if smart devices on your home network have been functioning as normal. A quick check has shown there have been no issues today, however you would like to check previous days.	
uc3	**sc30**		Concerned that unusual activity has occurred on the network in recent days, you would like to check the activity of one of your newly installed smart devices.	
uc4	**sc28**	You have established that unusual activity has occurred on the network, and that several smart devices have been active. You like to gather as much information as possible about a specific event to identify which smart device may have been compromised.		
uc5	**sc31**		You have identified that unusual activity occurred on your network in recent days, and would like to check if activity for a specific event was normal or unusual.	
uc6	**sc27**	You suspect unusual activity has occurred on the network today, and would like to check if your newly installed Smart Doorbell and Smart Bulb have been compromised.		
uc7	**sc33**			Having checked the daily usage of your Smart Bulb to establish if the daily activity was consistent, you would now like to do the same for the whole network. You would like to establish how many total normal activities occurred on the network for the previous two days.
uc8	**sc32**			Your newly installed Smart Bulb and Smart Doorbell appear to have been functioning normally for several days, having previously experienced unusual activity. You would like to establish if their daily usage is consistent or fluctuates over a given period.
uc9	**sc34**			Detecting unusual activity in devices with high activity fluctuations can be a challenge. You would therefore, like to be able to identify devices at risk in the future by monitoring all Smart Devices over several days, to establish which devices experience such fluctuations.

Table D.5: Scenarios used in Study 4 Chapter 7

		Day 1	Day 2	Day 3
uc1	**usc35**	You have just switched to a new Internet Service Provider and as part of the process you have installed a new home router. You would like to check the network is functioning normally.		
uc2	**sc38**		Your Internet Service Provider has sent you a repacement home router, since the previous one appeared to be experiencing issues. You successfuly installed the replacement router three days ago, and would now like to check activity on your home network during this time.	
uc3	**sc39**		You have established that some smart devices have successfully connected to the replacement home router, but others have not. You would like to know if your Smart Doorbell has successfully connected, and if so what type of activity it had yesterday.	
uc4	**sc37**	Having established which smart devices have successfully connected and been active on the new network, you would like to know if any of them are responsible for the unusual activity which has occurred.		
uc5	**sc40**		Despite installing the replacement home router, unusual activity has occurred on the network again. You would like to find out more information about the unusual activity, specifically what type of data communication it was.	
uc6	**sc36**	Your newly established home network appears to have experienced unusual activity. You are not sure which smart devices have successfully connected to the new network or if any devices are possibly compromised.		
uc7	**sc42**			Following a period of monitoring, you have established that the Smart Bulb is regularly generating more than the previously identified baseline of 100 normal activities per day. You now wonder if this may be true of other/all smart devices, suggesting a general and justified increase in normal daily activity.
uc8	**sc41**			You have established that on average your smart devices create roughly 100 normal activities each on your home network per day. You would now like to know if the activity for your newly installed Smart Doorbell has been consistent with this or has been unusually high/low.
uc9	**sc43**			You have been away for a weekend break, and have been using your Smart Camera to monitor your cat at home. You therefore, expected the camera to be one of the most active smart devices on your network, whilst you were away. However, since returning you have noticed the camera has continued to experience high activity levels, and you now wonder if you have possibly put the camera at risk by accessing it from the hotel wifi.

Appendix E

Thematic Coding Tables (Chapter 6)

The coding tables produced from qualitative analysis of responses in Chapter 6 are presented below. Responses were examined to identify common themes, ideas and patterns of meaning that came up repeatedly within the text. Example responses are provided for each Theme/Sub-theme.

Table E.1: **Aural** Agent Most Liked Features

	Theme	Sub-theme	Example Comment
Most Like	Usability	Convenience	"lots of people already own an Alexa so this would be a good way to get people to monitor their smart devices"
		Hands free	"I liked how it was hands free and didn't require a laptop etc"
		Quick	"much quicker than checking each device individually"
		Easy to Use	"easy to get updates about devices"
		Shared Responsibility	"how everybody in the home could share in monitoring their own devices"
	Accessibility	Visually Impaired	"it would be great for anyone visually impaired"
	Interactive	Enjoyable & Fun	"I enjoyed using this technology"
		Educational	"I liked learning new technology"
		New Experience	"I liked playing with an Alexa for the first time"
		Digital Assistant	"I like the idea of making better use of my my Alexa to assist me with other tasks rather than just listening to music"
	Awareness	Encouraged better Security	"This would actually convince me to care more about security"
		Improved Security	"non technical people like me can understand it"

Q. In a few words, what did you like most about the Aural Agent ? Why ?

Table E.2: **Aural** Agent Least Liked Features

	Theme	**Sub-theme**	**Example Comments**
Least Like	Usability	Laggy	"Was a bit laggy for a while"
		Voice recognition	"sometimes struggled to understand me first time"
		Bugs	"it crashed on me when using it"
	Functionality	Guidance	"I couldn't remember what each query did" "
	Design	Voice	"the Alexa should be customisable i.e. different voices"
		Query Design	"questions seemed a little long winded"
		Information	"I found it hard to remember some of the information if the answer was long"
	Privacy	Data Breach	"No sure I would want an Alexa in my home listening to my conversations"
		Trust	"Not sure I trust amazon with my data"
	Nothing	-	"nothing it was great"
	Unsure	-	"cant think of anything"

Q. In a few words, what did you like least about the Aural Agent ? Why ?

Table E.3: **Verbal** Agent Most Liked Features

	Theme	Sub-theme	Example Comment
Most Like	Usability	Convenience	"that I can check my devices are ok from my phone anytime I want"
		Quick	"Much quicker than opening my laptop to use some software"
		Easy to Use	"I found it easier to read some of the information on the phone"
	Accessibility	No extra device	"I like how this just runs on a phone so I don't need to buy another device"
		Portability	"I think most people would find the ability to use it anywhere the best feature"
	Interactive	Enjoyable/Fun	"Again it was actually cool to be able to chat with the bot and get the information"
	Design	Colours	"I liked the dark background theme"
	Awareness	Encouraged better Security	"This would actually convince me to care more about security"
		Improved Understanding	"I like how it simplifies something that would normally be quite difficult"

Q. In a few words, what did you like most about the Verbal Agent ? Why ?

Verbal Agent Least Liked Features

	Theme	Sub-theme	Example Comment
Least Like	Usability	Typing	"it was slow to type the questions out"
		Human Error	"typing each question could be prone to error"
		Slow	"it was slower than using the Alexa"
	Accessibility	Portability	"I never take my iPad outside the house so would be better on a phone"
	Functionality	Guidance	"lack of help to know what queries were available"
	Design	Query Design	"It was hard to figure out which query to use"
		Colours	"background was too dark"
	Nothing	-	"nothing it's a good idea"
	Unsure	-	"Don't know"

Q. In a few words, what did you like least about the Verbal Agent ? Why ?

Table E.4: **Aural** Agent Suggested Improvements

	Theme	**Sub-theme**	**Example Comments**
Improvements	Usability	Voice Recognition	"improve how it recognises peoples voice"
	Functionality	More queries	"quick summary for multiple days"
		Alerts	"alexa should tell you as soon as you say Hi that there have been issues on the network"
		Advice	"have alexa tell you what to do when you find unusual activity"
		Turn off Device	"have alexa disable infected devices"
	Accessibility	Portability	"not sure if this is possible but be able to speak to the alexa device from outside your house"
	Design	Voice	"everything works well but for fun add some celebrity voices e.g. Ricky Gervais the Office"
		Notification	"use the blue light on top of the echo when an attack occurs so when you walk in the room you know straight away to check your devices""
		Queries	"two of the questions gave too much information, this could be cut down to just the essential information"
	Nothing	-	"nothing I think it works well"
	Unsure	-	"can't think of anything"

Q. If you could suggest one improvement to the Aural agent what would it be ?

Table E.5: **Verbal** Agent Suggested Improvements

	Theme	**Sub-theme**	**Example Comments**
Improvements	Usability	Typing	"predictive typing like when you are texting on your phone"
		Command List	"some kind of reminder of the questions that are available"
	Functionality	More queries	"option to see which devices have been used most each day"
		Alerts	"alerts e.g. Your network has been hacked"
		Advice	"advice about what to do when my device is infected"
		Turn off Device	"ability to switch off a device which you think might be malfunctioning"
	Design	Colour	"different colours. I struggled to read the text on the blue background"
		Emojis	"I really liked it - maybe some cool emojis"
		Queries	"add the option of a weekend summary"
	Platform	Social Media	"never heard of Telegram, more people might use it if it was on something like Facebook messenger"
		Messenger	"it would be nice to run this in WhatsApp rather than having to download another app"
	Nothing	-	"nothing, seems pretty simple"
	Unsure	-	"can't think of anything"
	Categorised	Same	"same as alexa"

Q. If you could suggest one improvement to the Verbal agent what would it be ?

Table E.6: Device and Network Monitoring

	Theme	Sub-theme	Example Comments
Yes	Smart Device	-	"I routinely switch devices on to check them"
	Local Software	-	"sometimes scan my network using Kali"
	Network Device	-	"When I have time I occasionally check log files on my router"
No	Apathy	Time Consuming	"it's the kind of thing I put off then forget"
		Data not important	"didn't think anyone would be interested in my data"
		Risk/Reward Ratio	"the risk is quite low so I tend not to worry about it"
		Cost	"I don't want to buy more software"
		Unconcerned	"never been that concerned about my devices being hacked before"
	Awareness	Perceived Risk	"I didn't know my devices were at risk"
		Perceived Vulnerabilities	"I didn't think my smart devices were that vulnerable"
	Knowledge	Difficulty	"I'm not very technical so would struggle"
		Don't know how	"I don't know how to do it to be honest"
	Not considered	-	"I have never given it any thought"
	No smart device	-	"I do not own any smart devices"
	Uncategorised	No reason given	"No"

Q. Do you currently monitor your smart device activity or home network ?

Table E.7: Likeliness of Smart Device being Compromised

	Theme	**Sub-theme**	**Example Comments**
Yes	Awareness	Perceived Risk	"now that I know how hackers can use devices, I want to make sure they don't do it with mine"
		Perceived Vulnerabilities	"I didn't know how easy it was for devices to be hacked"
		Concerned	"it wasn't really something I was concerned about, but I have changed my mind""
	Knowledge	Difficulty	"I think so because it doesn't look quite as daunting as I thought"
		Don't know how	"it seems quite likely my devices could be targeted, so I should try to learn how"
	Uncategorised	No reason given	"yes"
No	Apathy	Time Consuming	"I should but I still think I would keep putting it off"
		Risk/Reward Ratio	"it's probably going to take one of my devices being hacked to make me start"
	No smart device	-	"I do not own any smart devices"
	Uncategorised	No reason given	"no"

Q. Do you think the likeliness of your devices being compromised is sufficient enough for you to consider starting ?

Table E.8: Likeliness of monitoring Smart Devices and Home Network with **Aural** Agent

	Theme	**Sub-theme**	**Example Comments**
Yes	Usability	Convenience	"I liked the convenience of using the Alexa"
		Quick	"I have no excuse as it's quick and I have an Alexa already"
		Easy to Use	"both Alexa and Telegram were easy to use"
	Awareness	Encouraged better security	"without the Alexa I probably wouldn't bother to monitor devices"
		Improved Understanding	"they would give me a better awareness of what was going on on my network"
	Uncategorised	No reason given	"yes"
No	Apathy	Time Consuming	"I know I should, but I probably wont find time to monitor my network"
		Data not important	"to be honest im still not sure I have any important data that hackers would want"
		Risk/Reward Ratio	"I don't think the risk is high enough to force me into action or warrant buying an Alexa"

Q. If you had access to the aural conversational agent at home would you be more likely to monitor your smart devices and home network ?

Table E.9: Likeliness of monitoring Smart Devices and Home Network with **Verbal** Agent

	Theme	**Sub-theme**	**Example Comments**
Yes	Usability	Convenience	"being able to check my devices anytime and anywhere using my phone would be very handy"
		Quick	"it looks quite simple and quick, so I would be more likely to check them"
		Easy to Use	"because without an app like the one I tested it would not be easy to keep an eye on devices"
	Awareness	Encouraged better security	"having the app on my phone I would have no excuse"
		Improved Understanding	"the tools made it easier to understand what my devices were doing, so I would be more likely to monitor them"
	Privacy	Data Breach	"I likely wouldn't use the Alexa for fear of eavesdropping, but Telegram yes"
		Trust	"I don't own an echo device because I don't like the idea of a device listening to my conversations. The telegram app would encourage me to monitor my devices more"
	Uncategorised	No reason given	"yes"
No	Apathy	Time Consuming	"I know I should, but I probably wont find time to monitor my network"
		Data not important	"to be honest im still not sure I have any important data that hackers would want"

Q. If you had access to the Verbal conversational agent at home would you be more likely to monitor your smart devices and home network ?

Table E.10: Ability to detect unusual Smart Device activity in the Future using the **Aural** Agent

	Theme	**Sub-theme**	**Example Comments**
Yes	Usability	Convenience	"I would be more likely to take care of my devices so would be able to spot any problems"
	Awareness	Encouraged better security	"it is easy to see each devices activity so I would know when a device is not functioning normally"
		Improved Understanding	"Alexa makes it easier to find out if unusual activity has occurred, and then find out exactly which device was at fault"
	Uncategorised	No reason given	"yes"
No	Apathy	Time Consuming	"again still seems like a time consuming job to stay on top"

Q. If you had access to the Aural conversational agent at home would you feel better equipped to detect unusual smart device activity in the future ?

Table E.11: Ability to detect unusual Smart Device activity in the Future using the **Verbal** Agent

	Theme	**Sub-theme**	**Example Comments**
Yes	Usability	Convenience	"I wouldn't be tied to only checking when in the house, so if I suspected something had happened I can check straight away"
	Awareness	Encouraged better security	"If I did decide to the use the Telegram chatbot I could check my devices a lot more, even whilst at work or on the bus"
		Improved Understanding	"Getting a quick summary of what devices were doing made it easier to monitor them, and know if they were doing something unusual"
	Uncategorised	No reason given	"yes"
No	Apathy	Time Consuming	"again still seems like a time consuming job to stay on top"

Q. If you had access to the Verbal conversational agent at home would you feel better equipped to detect unusual smart device activity in the future ?

Appendix F

Informed Consent (Chapter 7)

The study agreement and informed consent form is presented.

You are invited to participate in a research study titled Appreciation of Smart Device Activity. This study is being undertaken by **Christopher D. McDermott** from Robert Gordon University, Aberdeen.

The purpose of this research study is to measure your understanding and appreciation of threats facing smart devices in the home. The study will run for a **21-day duration**, and your participation in this study is entirely voluntary. You are free to withdraw at any time or omit any question.

To the best of our knowledge, no personally identifiable data will be collected during this study. We also believe there are no known risks associated with this research; however, as with any online related activity, the risk of a breach is always possible. To the best of our ability your answers in this study will remain confidential but may be used in future thesis and research paper publications. We will endeavour to minimise any risks by securely storing this data until the end of the research period, whereupon it will be destroyed.

If you agree to voluntarily engage in this research, and allow us to process your data in line with the University's privacy policy, please click the Next button below to give your informed consent, and start the study.

Appendix G

Usability Questionnaire (SUS) (Chapter 7)

The questionnaire used to assess the usability of the **Aural** and **Verbal** conversational agents presented in Chapter 7 is presented below.

Table G.1: System Usability Scale (SUS) Scores

	Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
I think that I would like to use this system frequently.					
I found the system unnecessarily complex.					
I thought the system was easy to use.					
I think that I would need the support of a technical person to be able to use this system.					
I found the various functions in this system were well integrated.					
I thought there was too much inconsistency in this system.					
I would imagine that most people would learn to use this system very quickly.					
I found the system very cumbersome to use.					
I felt very confident using the system.					
I needed to learn a lot of things before I could get going with this system.					

Appendix H

Chapter 7 Study Questions

The questions for each of the four studies in Chapter 7 are presented.

Table H.1: Study 1 Questions

#	**Day 1**
1	Has unusual activity occurred on the network today ?
2	Which source device had the first unusual activity today ?
3	Which of the following smart devices have used the network today ?
#	**Day 2**
1	Was any unusual activity detected on the network yesterday ?
2	How many normal and unusual activites did the smart camera have yesterday ?
3	Was activity 30 identified as normal or unusual ?
#	**Day 3**
1	What was the total number of normal activities for the whole network on each of the previous two days ?
2	How many Smart Camera activities were identified as normal on each of the previous two days ?
3	Which three devices have been most active for the five day period ?

Use-Cases are described in Tables 3.1-3.2

Study 1 Scenarios are described in Appendix D.2

Table H.2: Study 2 Questions

#	**Day 1**
1	Has there been any issues detected on the network today ?
2	What was the destination device of the first unusual activity today ?
3	Which smart devices have been active on the network today ?
#	**Day 2**
1	Did any unusual activity occur on the network yesterday ?
2	What percentage of normal and unusual activity did the Smart Camera have yesterday ?
3	Was activity ID 423 normal or unusual ?
#	**Day 3**
1	How many normal activities was there for the whole network on each of the previous two days ?
2	How many normal activities did the Smart Camera have on each of the previous two days ?
3	Which smart devices had the most activity yesterday ?

Use-Cases are described in Tables 3.1-3.2

Study 2 Scenarios are described in Appendix D.3

Table H.3: Study 3 Questions

#	**Day 1**
1	Has the network experienced any issues today ?
2	What was the ID number of the first unusual activity today ?
3	Which of the newly installed smart devices (Smart Doorbell or Smart Bulb) have been active on the network today ?
#	**Day 2**
1	Did the network experience any unusual activity yesterday ?
2	If the Smart Bulb was active on the network yesterday, what percentage of normal and unusual activity did it have ?
3	Was the smart device communication for activity ID 657 normal or unusual ?
#	**Day 3**
1	How many activities on the network were identified as normal on each of the previous two days ?
2	What was the number of normal activities generated by the Smart Bulb on each of the previous two days ?
3	Which three devices were most active on the network two days ago ?

Use-Cases are described in Tables 3.1-3.2

Study 3 Scenarios are described in Appendix D.4

Table H.4: Study 4 Questions

#	**Day 1**
1	Have any activities been identified as unusual on the network today ?
2	Which smart device was the source of the first unusual activity today ?
3	Which smart devices have successfully connected to the new network and had activity detected today ?
#	**Day 2**
1	Was all activity on the network yesterday identified as normal ?
2	How much normal and unusual activity did the Smart Doorbell have yesterday ?
3	What was the data type of activity ID 500 ?
#	**Day 3**
1	How many normal and unusual activities were there for the whole network, on each of the previous two days ?
2	How many normal and unusual of the previous two days ?
3	Has the Smart Camera been one of the three most active devices for either of the previous two days ?

Use-Cases are described in Tables 3.1-3.2

Study 4 Scenarios are described in Appendix D.5

Appendix I

Pre-Study/Post-Study Questionnaires (Chapters 6 and 7)

Pre-Study / Post-Study questionnaires are presented below.

Chapter 6

Table I.1: Pre-Study/Post-Study Survey

	Demographics (Pre-Study)
1	Please enter your age below *under 18* *18-24* *25-39* *40-59* *60+*
2	Please indiciate your level of technical knowledge, when using computers/smart devices *Novice* *Intermediate* *Advanced* *Expert*
3	Please specify if you own any of the following Smart devices *Amazon Echo/Alexa* *Google Home* *Smart Light Bulb* *Smart IP Camera* *Smart Thermostat* *Smart Door Bell* *Other*
	SA Statement (Pre-Study/Post-Study)
pe1	I am confident I can tell which smart devices are using my home network.
±	Smart devices are more secure than non smart equivalent devices.
pe2	I am confident I can tell how often a smart device is communicating on my homework, and how much of the available network bandwidth it is using.

±	Smart devices update themselves automatically.
pe3	I am confident I can tell which smart devices have the highest usage on my home network.
±	Smart devices are intelligent and can protect themselves from attackers.
co1	I am confident I can tell if my network is experiencing a normal level of device communications and bandwidth usage.
±	Smart devices alert you if an attacker is trying to compromise the device.
co2	I am confident I can tell if a smart device is functioning normally.
±	Smart devices are less likely to be targeted by attackers.
co3	I am confident I can tell if a smart device is using my home network more or less than normal.
pr1	I am confident I can tell if an attack has taken place on my home network.
±	Smart devices in the home are not accessible from the Internet.
pr2	I am confident I can tell if a smart device on my home network has been compromised.
±	Smart devices in the home can be used to perform attacks on the internet.
pr3	I am confident I could tell in the future if my home network or smart device had been compromised.
	Prototype Feedback (Post-Study)
1	In a few words, what did you like most and least about the **Aural** agent and **Verbal** agent ? Why ?
2	If you could suggest one improvement to the **Aural** agent and/or **Verbal** agent, what would it be ?
	Smart Device and Network Monitoring (Post-Study)
1	Do you currently monitor your smart device activity or home network ? *Yes* *No*
2	If you answered No above, do you think the likeliness of your devices being compromised is sufficient enough for you to consider starting ? *Yes* *No*
3	If you had access to the **Aural** agent and/or **Verbal** agent at home, would you be more likely to monitor your smart devices and home network ? *Yes* *No*
4	If you had access to the **Aural** agent or **Verbal** agent at home, would you feel better equipped to detect unusual smart device activity in the future ? *Yes* *No*

Questions highlighted Blue answered by Intervention group only

± used as a distractor statement

Chapter 7

Table I.2: Pre-study/Post-study Survey

	Demographics (Pre-Study)
1	Please enter your age below *under 18* *18-24* *25-39* *40-59* *60+*
2	Do you have a medically diagnosed hearing impairment ? If so, please briefly describe the impairment. *Yes* *No*
3	Do you have any medically diagnosed visual impairments (including colour-blindness) ? If so, please briefly describe them. *Yes* *No*
4	Do you have any medically diagnosed learning difficulties that might be relevant to this study (e.g. Dyslexia) ? If so, please briefly describe them. *Yes* *No*
5	Please indicate your level of technical knowledge, when using computers/smart devices *Novice* *Intermediate* *Advanced* *Expert*
6	Please specify if you own any of the following Smart devices *Amazon Echo/Alexa* *Google Home* *Smart Light Bulb* *Smart IP Camera* *Smart Thermostat* *Smart Door Bell* *Other*
7	Please indicate if you have used any of the tools used in the study before. *Amazon Alexa* *Telegram Messaging App*
8	If you own a smart device(s), do you currently monitor their activity on your home network ? *Yes* *No*
	SA Statement (Pre-Study/Post-Study)
pe1	I am confident I can tell which smart devices are using my home network.
±	Smart devices are more secure than non smart equivalent devices.
pe2	I am confident I can tell how much a smart device is using my home network.
±	Smart devices update themselves automatically.
pe3	I am confident I can tell which smart devices have the highest usage on my home network.
±	Smart devices are intelligent and can protect themselves from attackers.
co1	I am confident I can tell if my home network is functioning normally.
±	Smart devices alert you if an attacker is trying to compromise the device.

co2	I am confident I can tell if a smart device is functioning normally.
±	Smart devices are less likely to be targeted by attackers.
co3	I am confident I can tell if a smart device is using my home network more or less than normal.
pr1	I am confident I can tell if an attack has taken place on my home network.
±	Smart devices in the home are not accessible from the Internet.
pr2	I am confident I can tell if a smart device on my home network has been compromised.
±	Smart devices in the home can be used to perform attacks on the internet.
pr3	I am confident I could tell in the future if my home network or smart device had been compromised.
	Smart Device Security (Post-Study)
1	How realistic do you think it is that smart devices could be compromised and used as described in this study ? *Very Unlikely Unlikely Neutral Likely Very Likely*
2	Do you feel the risk of smart devices being compromised is high enough to justify the effort required to monitor them ? Why ? *Yes No*
	Learning Style (Post-Study)
1	In which way do you believe you learn most effectively ? *Visually Aurally (hearing/speaking) Verbally (reading/writing) Kinesthetically (by doing) Uncertain*
2	If you selected more than one learning style in the previous question, please rank you preferred mode of learning, in order of preference.? *Visually Aurally Verbally Kinesthetically*

Questions highlighted Blue answered by Intervention group only

± used as a distractor statement

Appendix J

Post-Study Interviews Questions (Chapter 7)

The questions asked in the post study interviews, carried out in Chapter 7, are presented below.

General

1. What was your overall experience of using the **Aural** and **Verbal** conversational agents ?

Perception

2. Please describe your experience of using the **Aural** and **Verbal** conversational agents to find out which devices were using the network and how much activity they had.

3. Did you prefer using the conversational agents or the visualisation tool to gather this kind of information ? Why ?

Comprehension

4. Please describe your experience of using the **Aural** and **Verbal** conversational agents to work out if the network or smart device was working normally. For example, discovering if any unusual activity had occurred or if a device was using the network more or less than normal.

5. Did you prefer using the conversational agents or the visualisation tool to gather this kind of information ? Why ?

Projection

6. Please describe your experience of using the **Aural** and **Verbal** conversational agents to determine if a smart device had been compromised, or an attack had taken place on the network.

7. Did you prefer using the conversational agents or the visualisation tool to gather this kind of information ? Why ?

Situational Awareness

When using the conversational agents did you feel they improved your:

8. Awareness about how smart devices were using the network. Why?

9. Understanding if a device's activity was suspicious or had changed. Why?

10. Ability to identify now or in the future if a device had been compromised or an attack had taken place on your network. Why?

11. In the Post Study survey, you said that it was *(very unlikely/unlikely/likely/very likely)* that smart devices would be compromised and used as described in this study? Can you explain why?

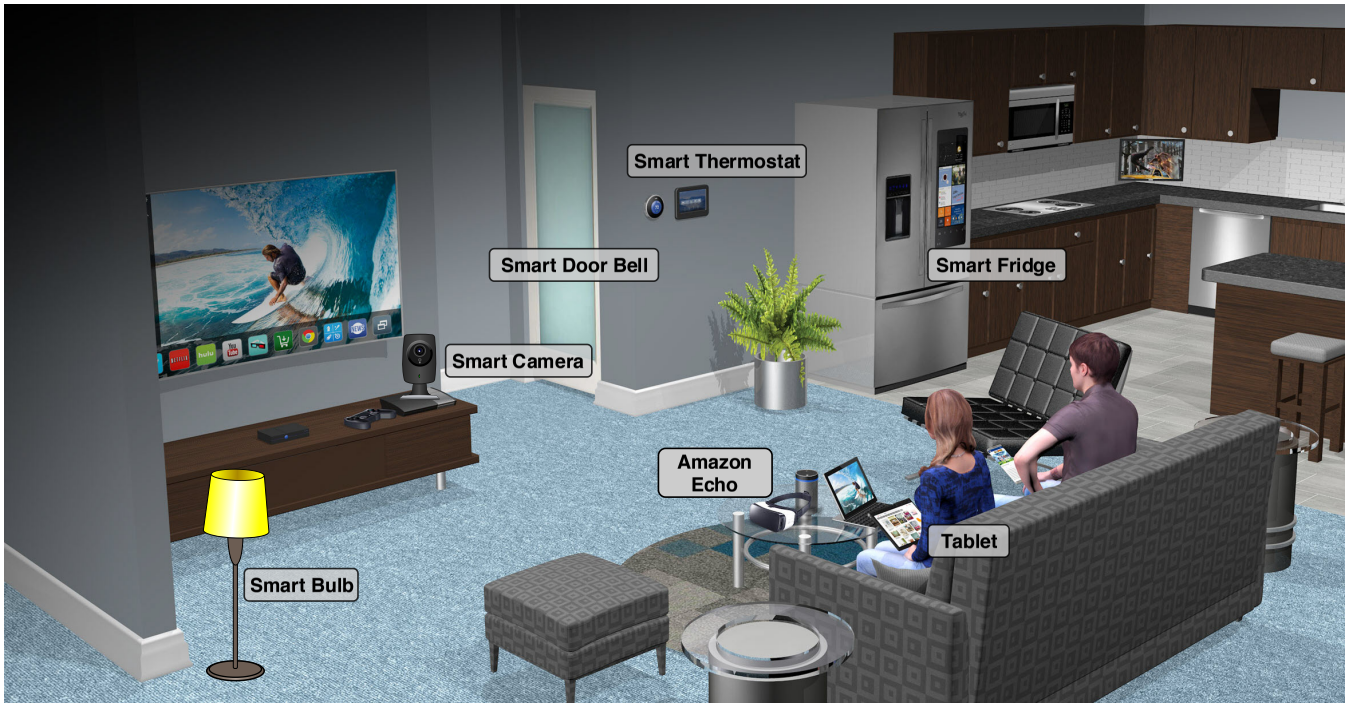
12. In the Post Study survey, you said you *(did/didn't)* feel the risk of smart devices being compromised is high enough to justify the effort required to monitor them? Can you explain why?

13. If you had easy access to the conversational agents at home do you think you would be more likely to monitor smart device activity? Why?

Appendix K

Smart Home Setup

The fictitious Smart home environment used in Chapters 4 - 7 is presented below.



Smart Home Environment