Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

Lloyd Nicolay Gustavsson

Assessing User Privacy in Virtual Assistant Devices via Passive Eavesdropping

Master's thesis in Information Security
Supervisor: Jia-Chun Lin
Co-supervisor: Ming-Chang Lee
June 2023

NTNU
Norwegian University of
Science and Technology

Lloyd Nicolay Gustavsson

Assessing User Privacy in Virtual Assistant Devices via Passive Eavesdropping

Master's thesis in Information Security
Supervisor: Jia-Chun Lin
Co-supervisor: Ming-Chang Lee
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

Assessing User Privacy in Virtual Assistant Devices via Passive Eavesdropping

Lloyd Nicolay Gustavsson

Abstract

This research investigated whether passive eavesdropping on virtual assistant devices would reveal any concerns for the privacy of users from engaging in interaction with these type of devices. The methodology used in this research assumed a predominantly quantitative approach. Numbers and measurable values were collected and used to make assumptions about privacy in connection with the use of virtual assistant devices. It also addressed the literature that is available on the part of privacy in virtual assistant devices, to further highlight and paint a picture of the broad privacy concerns that virtual assistant devices are still facing. These devices continue to be a cause of concern related to privacy because they are designed to be quick and easy to use for the individuals who use them. The devices can act as a source of increased productivity, but they are also widely used for leisure and entertainment. Often times they are used in households and workplaces, with many different users interacting with the same devices. Different users interacting with the same device can exacerbate the concerns around privacy. To facilitate easy interaction between the device and the users, manufacturers may decide on relaxing their efforts focused on privacy and security, and instead shift their focus to increase usability and interactivity. This put users of virtual assistant devices at risk of having their privacy violated. Some users may not express concerns about having their privacy violated, but it remains important to address the notion of privacy.

Sammendrag

I denne masteroppgaven ble det undersøkt hvorvidt passiv avlytting av nettverkstrafikk til og fra virtuelle assistentenheter ville kunne avsløre eventuelle brudd på personvern hos sluttbrukere. Metoden som er brukt i masteroppgaven har i all hovedsak vært en kvantitativ tilnærming. Tall og statistikk som kan måles, ble samlet inn over en lengre periode for å undersøke forhold rundt personvernet i forbindelse med bruk av enhetene. Tilgjengelig litteratur som omhandler personvern i sammenheng med bruk av virtuelle assistentenheter, ble også addressert for ytterligere å belyse eventuelle brudd på personvern som kan oppstå i forbindelse med bruk av virtuelle assistentenheter. Disse enhetene er til stadighet forbundet med uro angående personvern fordi de er utformet for å være tilgjengelige, raske og enkle å bruke for sluttbrukere. Enhetene kan føre til økt produktivitet, men de er også mye brukt til fritid og underholdning. Derfor brukes de ofte av flere ulike brukere i for eksempel en husholdning eller på en arbeidsplass. Mange ulike brukere av disse assistentenhetene kan øke risikoen for at personvernet blir krenket. For å sørge for at enhetene fortsetter å være enkle å bruke, kan produsentene ofte se mellom fingrene hva angår personvern og sikkerhet, og i stedet flytte fokuset over på å øke brukervennlighet og interaktivitet for sluttbrukeren. Fra denne synsvinkelen fremgår det som kan bidra til at de som bruker disse enhetene har større risiko for å få krenket sitt personvern. Noen brukere uttrykker kanskje ikke bekymring for å få deres personvern krenket, men det er fortsatt viktig at det vektlegges å ivareta personvern så godt som mulig.

Contents

Figures

Tables

Acronyms

ASR Automatic Speech Recognition. 6

CSV comma-separated values. 16

IoT Internet of Things. 5, 6, 9, 10

IP internet protocol. 15

MAC Media Access Control. 34, 35

NLG Natural Language Generator. 6

NLP Natural Language Processing. 6

OSI Open Systems Interconnection. 35

WLAN wireless local area network. 12–14, 34

Glossary

4G Fourth generation cellular network technology. 13, 14, 25, 29

Wi-Fi Wireless network protocols. 2, 12, 13, 21, 35

Chapter 1

Introduction

This Chapter will in brief introduce key areas of the research such as the problem domain of the research to be carried out, the research questions, scope and delimitation, and the motivation behind performing the research. Lastly, it will detail the structure of the rest of the thesis.

Virtual assistant devices and its associated technology is on its path to become ubiquitous in consumer households around the world. The software that enables the implementation and functionality of virtual assistants, can now be found in a vast array of devices, and multiple environments. These environments include the kitchen at home with increasingly smarter appliances, the cars that people drive, and the smartphone that can be found in the pockets of many. Previously, the early forays into virtual assistant devices was limited to specific devices exhibiting interactive capabilities by voice with its user. In general, virtual assistants of today are now more often than not, bundled together with the operative systems as software that ships with the device. Virtual assistant software in the form of Siri from Apple, Cortana from Microsoft, Alexa from Amazon, and the Google Assistant from Google is deployed together with each vendor's platform and operative systems to fit in with a diverse range of devices.

1.1 Problem domain

Virtual assistant devices can provide accessible and intuitive functionality to its users. With this ease of use, however, concerns surrounding privacy and security quickly present themselves. Virtual assistant devices such as the Amazon Echo has been criticized for lacking solid means to provide for increased security and privacy on behalf of its users. First and foremost, these type of devices seek to make it as easy and as quick as possible for users to set up their device, and immediately start taking advantage of its features.

1.2 Research questions and objective

The objective or goal of this research is to more closely look at privacy and security-related issues in connection with virtual assistant devices. Furthermore, it will be of great interest to establish any worthwhile insight as to whether user presence can be passively inferred from interpreting the nature of any packets collected.

The research questions that were formulated to guide further research on the topic of virtual assistant devices:

- Can user presence be inferred from passive eavesdropping on virtual assistant device network traffic?
- What privacy and security-related challenges are virtual assistant devices facing?

1.3 Scope and delimitation

The scope of research question 1 will involve the exploratory, practical, and real-world part of the research. This investigation will be delimited to the nature of operations in the Wi-Fi protocol for both devices in question. The scope of the aforementioned research will entail investigating the transmission of network traffic that is passed to and from two virtual assistant devices in two different use-cases that will be described in more detail in the following chapters.

In the case of research question 2, a literature study is conducted. The scope of this study will limit itself to the state-of-the-art literature on the topic of privacy and security-related issues, which will be consulted to reach a consensus on the challenges that present themselves from using virtual assistant devices.

1.4 Thesis structure

The remainder of the thesis will encompass multiple Chapters. These are the background Chapter, a related work Chapter, the Chapter on methodology, an analysis Chapter, the discussion Chapter, and the conclusion Chapter.

The background will be presented in its own Chapter. It will shed some light on virtual assistant devices and the underlying technology on which it operates. It also will present the typical environments in which these type of devices are commonly used. The premise of this research is that the reader possess zero knowledge on this topic to begin with. It is desirable to be able to communicate these ideas to the reader in such a way that they can draw as much benefit as possible from reading the rest of the thesis.

The related work Chapter highlights available information pertaining to the research questions.

The methodology is next. This Chapter details the planned approach for collecting, interpreting, and disseminating information from relevant sources. It will

also lay out the topology of the suggested experimental setup for carrying out different scenarios.

In the analysis Chapter, an assessment of the collected data material will be conducted. Hopefully, answers will be found that can help to assist with refuting or accepting the research questions that were asked previously.

The discussion Chapter will revolve around to what degree the research questions have successfully been addressed. It will also discuss in what ways the research conducted is relevant to the field of study, and how it can help to assist with interpretation of the threats that are extant in a virtual assistant device environment.

The final part is the conclusion. This Chapter will revisit the most important contributions of the research yielded. It will attempt to describe the findings of the research, and communicate this to the reader in an understandable manner.

Chapter 2

Background

This Chapter will seek to communicate an overview of the environment that virtual assistant devices like the Google Nest and the Amazon Echo constitute a part of. It will also attempt to shed some light on the architecture of virtual assistant devices.

Virtual assistant devices are most commonly used within the Internet of Things (IoT) environment. The IoT environment has made an entrance into many different sectors, industries, and numerous other areas of society. IoT is often characterized by many smaller devices which exhibit a high degree of interoperability between each other. Since the inception of the IoT environment and associated technologies, traditionally stand-alone systems and networks have gradually seen a shift from being isolated to being interconnected with each other, and even acquire communication on a global scale. This new environment has brought forth renewed promises of increased usability and flexibility to meet a new technological reality. It has enabled for new proving grounds for innovative technology. While the possibilities has increased in the technological domain, both the consumers and the manufacturer are not fully aware of the extent of the risks and implications that may arise from using these devices. The focus of this research is centered in on the part of the IoT environment that surrounds the smart home. A smart home is a home that utilise the capabilities of IoT and its devices for intentions such as to automate tasks that had to be carried out manually in the past. It can also increase efficiency in ways such as for example spending shorter time on chores, or be able to have more work done than what was previously possible. Another large reason for the adoption of IoT devices is its ability to enable consumers quick access to vast amounts of information and entertainment in ways that were previously unprecedented.

As can be observed in Figure 2.1, the general architecture of a virtual assistant device span across four principal components. The first and most important component is the user, providing requests and input for the device and any associated companion apps. The second and third components are the virtual assistant devices and any corresponding companion apps. Each of these components can take input or requests from a user and communicate results back to the user. In

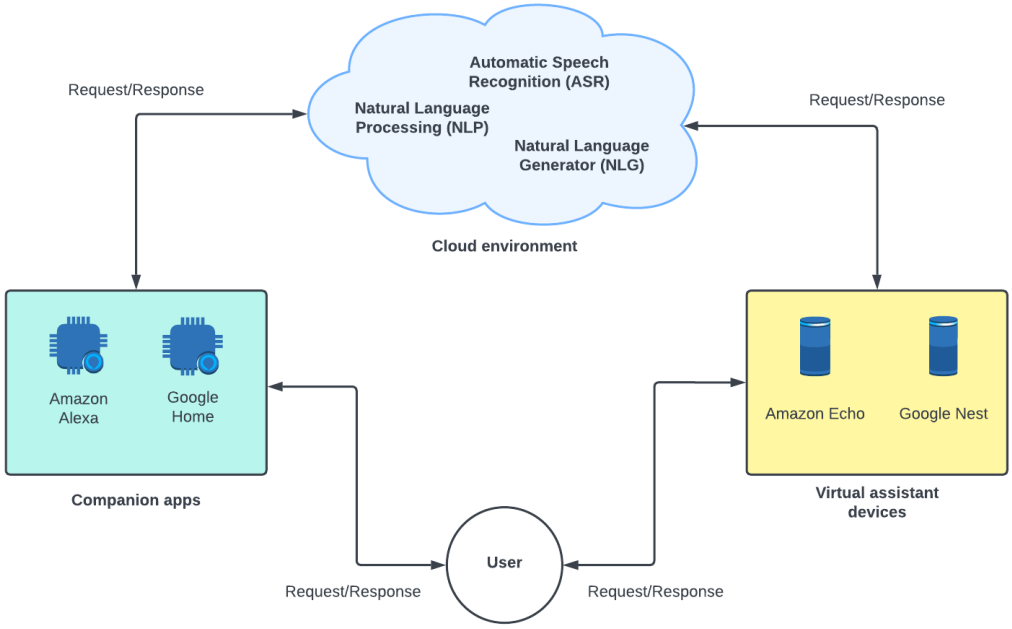


Figure 2.1: A general overview of the virtual assistant device architecture

a typical IoT environment, they can also send and receive information to simpler devices such as light bulbs and thermostats. The fourth and final component is the cloud environment. This is an often proprietary environment of the vendor which is largely inaccessible for any user of virtual assistant devices although mostly all of the data a user provide is sent to this component for processing. It is difficult for the average user to be certain about what happens to this data, where it is stored, or who may have access to it, such as third-party interests.

For the fourth component which is the cloud architecture, several processes, algorithms, or machine learning techniques are in place to interpret the requests and input from a user to the best of its ability. In the case of the user interacting with one of the virtual assistant devices used in the research, an audio recording is made. This audio recording is not interpreted in the device itself. Rather, it is transmitted to the cloud environment for further processing. When it arrives in the cloud environment, it is the responsibility of the Automatic Speech Recognition (ASR) process to properly translate any speech contained in the audio recording from speech to a text format [1]. Afterwards, Natural Language Processing (NLP) make an attempt to interpret the text provided. When NLP has interpreted the audio recording, the cloud environment may agree on a response which will be communicated back to the user [1]. In case any part of the response is to be communicated back to the user through speech via the speakers of the virtual assistant device, the text response must be synthesized back to speech in the form of a new audio recording. The Natural Language Generator (NLG) technique will

attempt to take care of this [1].

Chapter 3

Related work

This Chapter chronicle some of the practical and theoretical research that is available for addressing privacy in virtual assistant devices. It looks at two distinct sections, with each of them related to the research questions put forward in Chapter 1.

3.1 Passive eavesdropping

Gu et al. [2] provide a comprehensive investigation into inference of events across distinct smart home devices in an IoT environment such as a temperature and vibration sensor, a motion sensor, a sensor for detecting water leaks, and a smart light bulb. They managed to fingerprint the events for the aforementioned IoT devices. They also managed to identify inter-device chain of events across the devices whereas if for example the motion sensor detected movement, it would notify a hub in another event, which in turn notified the smart light bulb to turn itself on [2].

Apthorpe et al. [3] investigated four IoT devices with encrypted traffic. Devices included an Amazon Echo device, as well as a sleep monitor and a motion sensing camera. For interaction with the Amazon Echo device, the researchers asked the device about the weather, the time of day, and distance between locations. They showed how graphs of the network traffic to and from each device increased in accordance with user interaction.

R. Jackson et al. [4] achieved high accuracy in their endeavour to classify encrypted traffic transmitted between an Amazon Echo device and the cloud environment of Amazon. They employed six machine learning techniques on three different type of feature vectors made on the basis of tcptrace, histogram, or a combination of the two [4].

Dong et al. [5] used machine learning techniques to profile the type of devices from an IoT environment with high accuracy. The type of devices investigated included virtual assistant devices, smart bulbs, smart cameras, and more generic devices such as mobile phones and tablets.

In [6], the k-mean clustering machine learning technique was used to investigate if the microphone mute button worked as intended. It was also investigated if any trace of conversations was recorded and streamed in the absence of a wake-word being spoken. According to the results obtained, it was concluded that the functionality of the mute button was trustworthy. Unfortunately, on the part of privacy, the researchers found that the device had indeed recorded and streamed audio recordings to the cloud environment even when the wake-word was not used [6].

Trimananda et al. [7] presented PingPong, a software tool for establishing packet signatures in smart home devices. They used pair clustering to derive packet-level signatures for events. The events spanned many simpler IoT devices such as thermostats, door locks, cameras, and light bulbs that perform smaller, specific events. Their results showed that many packet-level signatures can be accurately obtained for the different devices, with a low rate of false positives [7].

A fingerprinting of voice commands is proposed in [8]. In this research, packet sizes, the number of packets, ingress and egress traffic, and durations, are all used to try and attribute fingerprints to different voice commands from encrypted packet traffic. Several neural network machine learning concepts were used to achieve the fingerprinting. One of the downsides of this research was that actual human voices was not utilized. Automated voices was used instead [8].

3.2 Challenges in security and privacy for virtual assistant devices

Cheng et al. [9] performed an extensive literature survey on security and privacy for voice assistant devices. Four categories were identified as possible stepping stones for launching more specific attacks. They found that access control could be circumvented for some of these type of devices. They also identified that the audio components in the devices such as the microphone can be subjected to acoustic attacks, and the properties of acoustics to be able to infer dimensions and spatial properties of the room in which the voice assistant device is placed [9].

Bolton et al. [10] investigated privacy in addition to security challenges in virtual assistants. After consulting numerous sources on security and privacy for virtual assistants, it was found that end users may not worry much at all about the privacy concerns that present themselves from using virtual assistants and virtual assistant devices. The researchers also came to the conclusion that the mechanisms that were in place for monitoring third-party applications or skills for the virtual assistants are lacking in making sure they do not carry malicious intent.

In [11], a DDoS attack was launched, which successfully severed the connection to the internet for a Amazon Echo Dot device by using a syn-flood attack.

It is apparent from the literature that virtual assistant devices face many challenges.

Chapter 4

Methodology

Devising a methodology for the purpose of conducting the research posed in this paper was a multi-stage process. Firstly, it had to be considered what scenarios would be viable and feasible for carrying out the research. Secondly, to be able to gain any insight into nuances of research question 1, it was deemed necessary to establish two slightly different real-world environments to replicate everyday interaction with the devices that was procured for the purpose of this research. The methodology is best described as quantitative in its approach as it is concerned with numbers and values that can be measured.

In Figure 4.1, an overview of the network topology is presented.

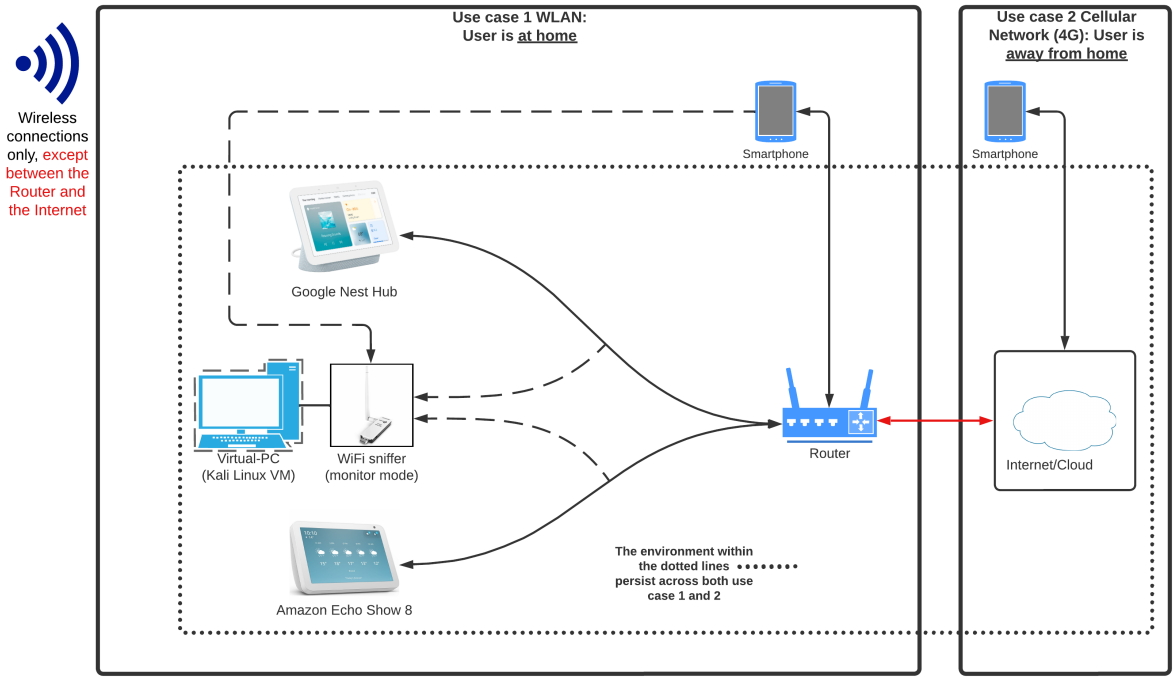


Figure 4.1: Diagram showing the topology of the network.

Solid lines in the Figure represents the regular channels for communication between devices. The dashed lines represent the interception of packets by the sniffing device. The environment encased by dotted lines indicate what part of the topology is preserved across the two use cases.

The two virtual assistant devices used for the research is the Google Nest Hub running the Fuschia OS operative system, and the Amazon Echo Show 8 running the Fire OS operative system, from the vendors Google and Amazon respectively. For the purpose of sniffing Wi-Fi packets, the TP-Link TL-WN722N Wireless USB Adapter was used. This particular model is one of few models available that supports monitor mode

The router for facilitating network traffic transmission to and from the wireless local area network (WLAN), and which functions as an access point for the other devices in the network topology, is a ZyxelFMG3542. The Virtual-PC is powered by the 2022.4 release of the Kali Linux operative system. The underlying physical hardware that enables the execution of the Virtual-PC is a desktop workstation running a Windows operative system. A Virtual-PC running a Linux distro was necessary because of Windows' inability to properly support monitor mode The smartphone used is the iPhone X running the software version iOS 15.3.1. To enable communication between the smartphone and the virtual assistant devices in use case 2, there is a companion app supplied by each vendor for facilitating communication from the smartphone. For the Amazon Echo Show 8, the corresponding companion app is the Amazon Alexa app, running on version 2.2.518779. For the Google Nest Hub, the corresponding companion app is the Google Home app, running on version 2.63. The application on the Virtual-PC that will facilitate capture and storage of the packets obtained through theWi-Fi sniffer, is an application known as Wireshark, often used for analysis of network protocols. The version of Wireshark used will be the 4.0.1 version.

For the research, it was decided to settle on not just one, but two virtual assistant devices. It would be interesting to investigate only one device in isolation. By using two virtual assistant devices, this leaves more room for making a comparison across two quite similar devices that offer many of the same functions. The devices also stem from the two largest vendors in the market for these type of devices. This enables the research to make somewhat of a comparison on the performance gap between these two vendors' virtual assistant devices. Naturally, this is not the primary effort of the research, and both vendors supplies the market with numerous other devices that fulfills much of the same needs for consumers. This means that any differences found in this research between the vendors products will likely not be of any empirical significance. However, the difference could be an interesting observation to point out for the reader.

Two different use cases can be observed in Figure 4.1. All communication will take place within the wireless domain with the focal point of the research being the 802.11 protocol, most frequently referred to as Wi-Fi in colloquial terms. Both the Amazon Echo Show 8 and the Google Nest Hub device supports Bluetooth in addition to Wi-Fi, but seeing as the primary means of communication to and from

these devices is conducted by utilizing Wi-Fi, Bluetooth is not considered at this point in time.

The first use case concerns having all devices associated with the research connected to the same WLAN. The second use case concerns having all devices associated with the research connected to the same WLAN. The exception is the smartphone which in this use case has been disconnected from the WLAN, and instead is connected to the global cellular network, also known as 4G. The assumption from connecting the smartphone to an entirely different network then, is that the observation of traffic in the second use case will manifest differently from the traffic observed in the first use case.

The motivation behind segmenting the two different use cases is to simulate whether a device associated with the virtual assistant devices can be inferred, or otherwise profiled, to be remote or on-site. On-site in this case refer to the party interacting with the device being at home. And remote refers to the party interacting with the device being away in that the party's smartphone is interacting with the device remotely with its traffic originating in a different network. It will be interesting to see if any discrepancies can be derived from interpreting the capture network traffic from each of these use cases respectively. If there are any meaningful results found, it could indicate the viability of an attacker to capture network traffic in the vicinity of a target household to infer if at least one party is away from their house.

Four different scenarios were devised as shown in Table 4.1. During the first scenario, the user is in the same room as the virtual assistant device, issuing voice commands. In this scenario, the virtual assistant device is in a muted state. The traffic or information obtained in this scenario will be a point of reference for the other scenarios, being the baseline data. Scenario two is similar to scenario one. The same voice commands shall be issued. But during this scenario, the virtual assistant device will assume an unmuted state. This will serve a two-fold purpose for the research. First, it will shed some light on whether there is any noticeable pattern of increased traffic when compared to the baseline data in the previous scenario. If this is not the case, it may be that the muted state for the device is not working as intended. In scenario three, the smartphone has become the primary means of executing the voice commands. However, the commands are not conveyed through voice as speaking to the companion app in the smartphone does not enable communication with the virtual assistant device. Rather, both companion apps for the virtual assistant devices from Google and Amazon respectively, introduce a functionality that is known as Routines. In this scenario, the virtual assistant device is set to a muted state similar to scenario one. The fourth and final scenario mirrors the third scenario, but the virtual assistant device is set to an unmuted state as was the case in scenario two.

Table 4.2 shows how scenario 1 and 2 is intended to be carried out with use case one, and with scenario 3 and 4 intended for use case two.

For use across all scenarios, a sequence of events was established. These events were chosen on the basis of what events the researcher found to be frequently

Table 4.1: Description of each scenario.

	Description
Scenario 1	Interact physically with the device, device is muted
Scenario 2	Interact physically with the device, device is unmuted
Scenario 3	Interact with the device remotely through a companion app, device is muted
Scenario 4	Interact with the device remotely through a companion app, device is unmuted

Table 4.2: The relation between use cases and scenarios.

Use case 1: WLAN		Use case 2: 4G	
Scenario 1	Scenario 2	Scenario 3	Scenario 4

used with the virtual assistant devices. It can also be argued that the events were chosen according to what type of events that could be assumed to be used quite frequently when the owner is absent from home and at activities such as going to work. The first event was defined as asking for the weather. For practical purposes, the location queried was chosen to be consistent across scenarios. Thus, the location was set to the city of New York. In event number two, asking the device for directions was of interest. Yet again, to keep the request consistent across scenarios, the destinations for the query was set to the cities of Frankfurt and Berlin in Germany. In the third event, an audio call is performed. This event is interesting to include as it differs the most from the other events. It is also of interest to use with a virtual assistant device as it can allow for accessible and easy communication with other members of a household with affiliated devices. In the last event, an announcement is made. The announcement is broadcast on the virtual assistant device regardless of the announcement being performed from the smartphone or the virtual assistant device itself. The different events is shown in Table 4.3.

Table 4.3: Description of each event.

	Description
Event 1	"What's the weather in New York?"
Event 2	"Give the directions from Berlin to Frankfurt"
Event 3	"Call [name of device]"
Event 4	"I leave for work"

For every scenario, five to ten captures will be performed. There is no specific motive behind the number of captures to perform other than the fact that it is desirable to perform captures until the values they produce shows a tendency to somewhat converge against similar values. Or that the captures display similar characteristics. Each scenario must also be performed for each of the virtual assistant devices. Furthermore, the capture is performed only for one of the virtual assistant devices at any given time. This is done to ensure that any inter-device communication between the two virtual assistant devices is kept at a minimum to more accurately represent the behavior of the device operating in a stand-alone

manner within a given household environment. It is also more reflective of the observation that the majority of households typically owns only one virtual assistant device [12]. While the capture is carried out for one virtual assistant device, the other device is powered off. Additionally, the captures are made quite some time after a virtual assistant device has been powered on. This is to ensure that the captured traffic is not influenced by initial delegation of internet protocol (IP) addresses, power management control, or other functionality that often present themselves when the device is just recently connected to a network. It is desirable to keep the device as dormant as possible before a capture procedure begins. All captures will be characterized by the sequence of events given in table 4.3. For each of the events, they will take place at specific points in time during a capture. The first event will be instantiated after 10 seconds. The second event comes at one minute or 60 seconds. The next event is started after two minutes or 120 seconds. Finally, event four is executed after four minutes or 240 seconds. The requests given to execute the events are given on a best-effort basis. Spoken requests meant for the virtual assistant devices in scenario 1 and 2 are given as close to the normative points in time as described for each event. The same principle is applied in scenario 3 and 4, where the requests are given from the companion app on the smartphone as precisely as possible. The total duration of each capture is five minutes or 300 seconds. The duration can accurately be configured in Wireshark before each capture.

4.1 Analysis flow

In Figure 4.2, the workflow associated with the concrete research is shown.

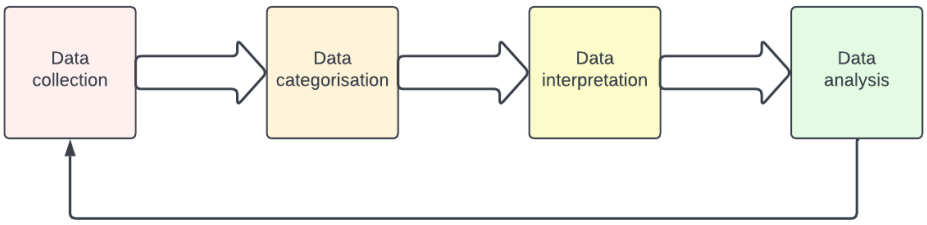


Figure 4.2: Diagram showing the general workflow for the analysis

The generalized approach to conducting the analysis will be to follow the steps as they are laid out in the Figure. Data collection will encompass the act of capturing packets. It is important to carry the capture out as correctly as possible. A correct capture involves keeping the timing of events consistent as best as possible, making sure the proper capture filters is applied in order to capture traffic for the intended devices. Another important aspect is to make sure there is little noise in the surrounding environment, seeing as the risk of inadvertently capturing other events is present.

Next is the categorisation of data. This involved labeling and naming the files from the data collection in relation to the device that it was intended to capture for. It also involved segmenting different parts of the captures according to whether packet traffic was outbound or inbound for the virtual assistant device. Another segmentation of the capture was to differentiate between the number of packets per second, and the number of bytes per second. The information segmented was further exported to files consisting of comma-separated values (CSV).

The third step was the interpretation of data. The CSV files obtained from the previous step was loaded into Microsoft Excel for further interpretation. At this stage, two approaches was of interest. One is to graph the data from the CSV files in a visual-based approach. The expectation is that one will be able to visually observe trends and patterns in the captured data. The other approach is to look at the statistics such as total number of packets or bytes, and the percentage shares between outbound and inbound data. This approach gives a more in-depth view of the actual numbers across the the capture as whole, but in comparison to the visual approach it may be less telling of patterns, especially in relation to the events.

The fourth and final step before backtracking to the start of the workflow in preparation for a new capture, is the data analysis. This step involves analysing and scrutinizing the graphs and statistics from the previous step. In this step, it is desirable to compare both the statistics and graphs between scenarios of interest. The comparisons will hopefully assist with answering the research questions that were posited in Chapter 1.

Chapter 5

Analysis

This Chapter will chronicle the investigation launched with an emphasis on interpreting the packet collection from the virtual assistant devices.

5.1 Scenario 1

In scenario 1, the Amazon Echo Show 8 and the Google Nest Hub were both set to a muted state as was described in Chapter 4, and the sequence of events are carried out with the devices through voice commands. The captured traffic in this scenario serves as the baseline traffic for traffic captured in the next scenarios.

Looking closer at the packet traffic for the Google Nest Hub in this scenario, there are multiple observations to be made. There is a consistently larger share of outbound packet traffic across all captures. Using the display filter "wlan.fc.type_subtype == 0x24 && wlan.sa == 1c:53:f9:bd:c4:f7" that combines the notion of outbound packets and null data or null function packets, shows that all null data packets are outbound packets. This suggests the device receives little data over the duration of each capture. Instead, it mostly spend its time on transmitting null data packets to the router in order to stay operational, connected and balanced in accordance with operative parameters for maintaining device functionality. Evidently, since all transmitted outbound data consists of null data packets that carries no data, the device did not record audio to be passed on to any cloud environment. Suggesting that the interaction and sequence of events with voice does not relay any traffic further on in the network topology than the device. Thus, it can be assumed that the device is not making an audio recording, or that it may record but is administered to not act on recorded data, or transmit it any further. The device in total send and receive 13785 packets on average across the combined captures.

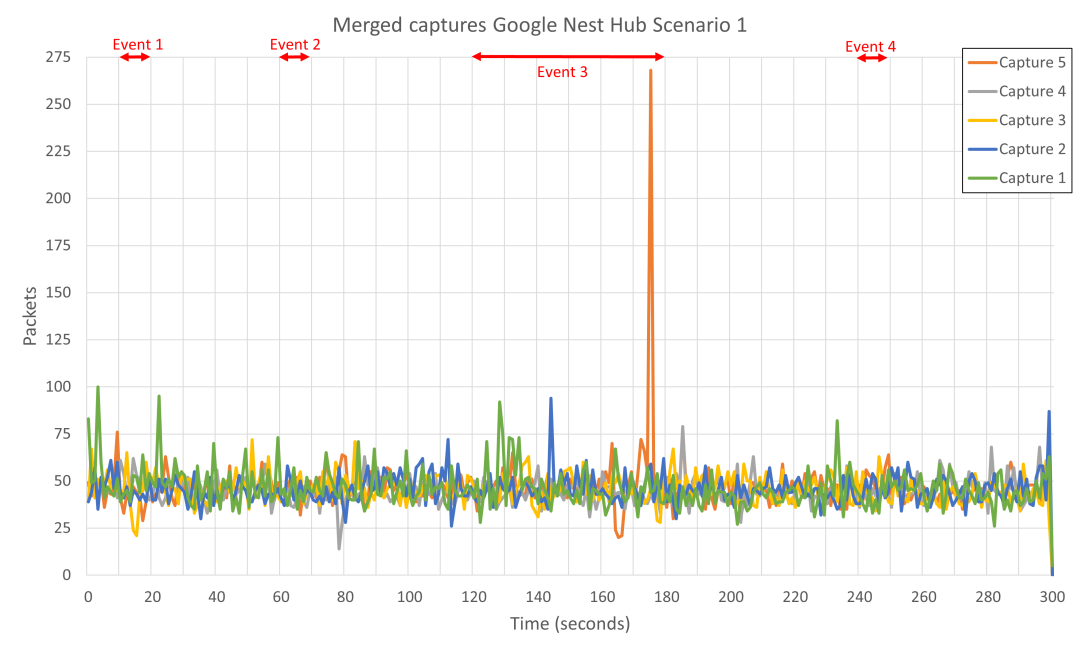


Figure 5.1: The total number of packets transmitted per second per capture session for the Google Nest Hub

Table 5.1: Statistics for the Google Nest Hub Scenario 1

Google Nest Hub Scenario 1					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	72,4 %	73,1 %	72,2 %	74,7 %	71,9 %
Non-null data bytes	58,5 %	56,5 %	55,4 %	52,7 %	61,3 %
Outbound packets	82,7 %	84,5 %	83,0 %	86,0 %	83,0 %
Outbound bytes	64,5 %	69,6 %	67,4 %	72,8 %	63,1 %
Inbound packets	17,3 %	15,5 %	17,0 %	14,0 %	17,0 %
Inbound bytes	35,5 %	30,1 %	32,6 %	27,2 %	36,9 %
Total # of packets	13910	13721	13735	13467	14090
Total # of bytes	1116822	1059498	1024288	979033	1204588

In Figure 5.2, the number of bytes per second is displayed. It doesn't reveal any new information, as it quite closely follows the pattern observed in 5.1. The advantage of displaying the byte traffic per second is that it is easier to use it for observing when actual information is transmitted compared to packet traffic per second.

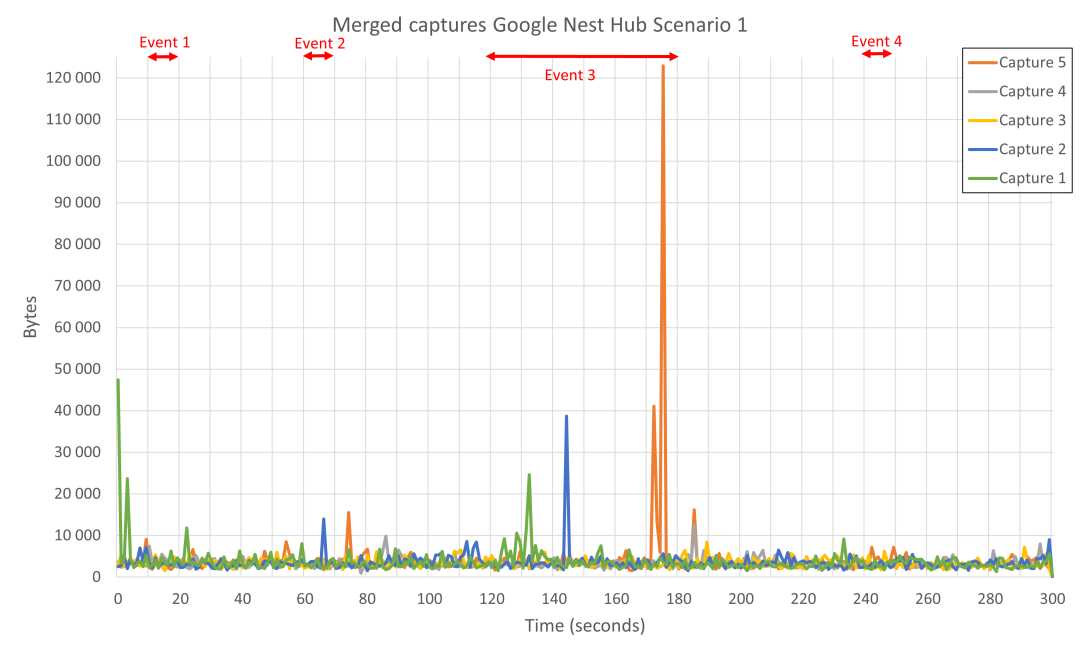


Figure 5.2: The total number of bytes transmitted per second per capture session for the Google Nest Hub

For the Amazon Echo Show 8, the packet traffic behaves differently than is the case with the Google Nest Hub. The frequency and total number of packet transmissions is quite low. On average, the device is only subject to a single packet per second. Similarly to the Google Nest Hub, applying the display filter "wlan.fc.type_subtype == 0x24 && wlan.sa == 44:6d:7f:b6:47:65" that combines the notion of outgoing packets and null data packets, shows that all null data packets are outbound packets. Compared to the Google Nest Hub, the number of null data packets is very low. The majority of traffic for the Amazon Echo Show 8 device is inbound, an inverse trend when compared to the inbound traffic of the Google Nest Hub device. It can also be observed in Figure 5.3 that every capture on the Amazon Echo Show 8 display at least one pronounced random spike in packet traffic during the capture, an observation which is less evident when looking at Google Nest Hub. The device in total send and receive 358 packets on average across all captures.

The byte traffic per second of the Amazon Echo show 8 as shown in Figure 5.4 closely follows the pattern and movement of the corresponding packet traffic as was shown in Figure 5.3.

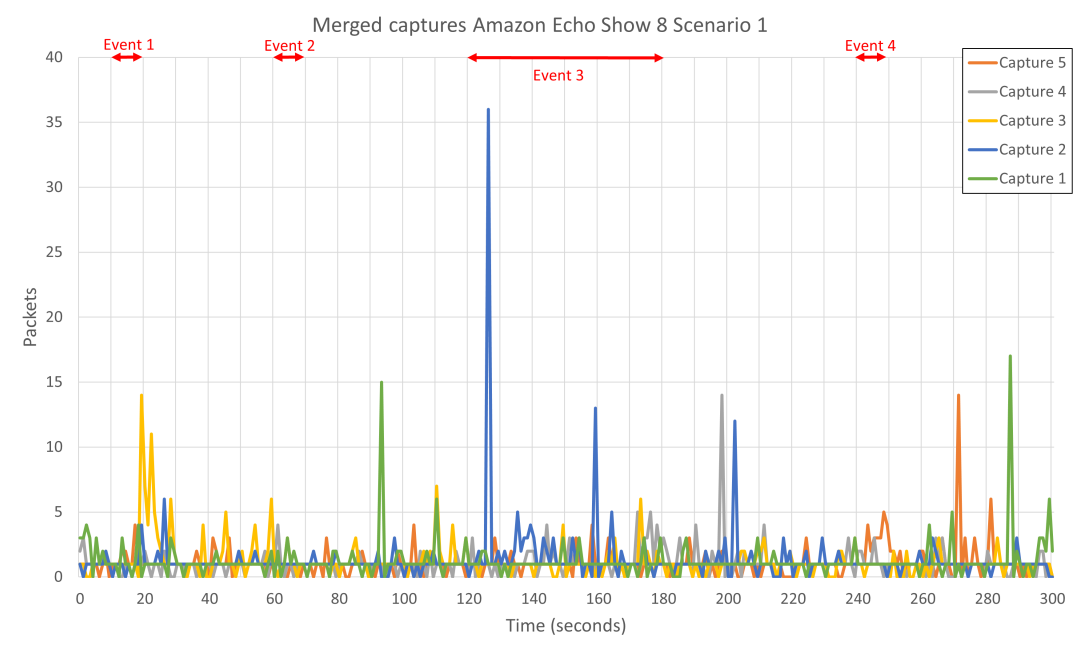


Figure 5.3: The total number of packets transmitted per second per capture session for the Amazon Echo Show 8

Table 5.2: Statistics for the Amazon Echo Show 8 Scenario 1

Amazon Echo Show 8 Scenario 1					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	1,3 %	0,8 %	1,9 %	2,8 %	1,2 %
Non-null data bytes	99,5 %	99,7 %	99,3 %	98,8 %	99,5 %
Outbound packets	14,2 %	14,9 %	18,3 %	17,7 %	14,2 %
Outbound bytes	25,6 %	36,5 %	31,7 %	36,0 %	33,4 %
Inbound packets	85,8 %	85,1 %	81,7 %	82,3 %	85,8 %
Inbound bytes	74,4 %	68,5 %	68,3 %	64,0 %	66,6 %
Total # of packets	373	397	360	327	331
Total # of bytes	44378	43159	48977	34419	33610

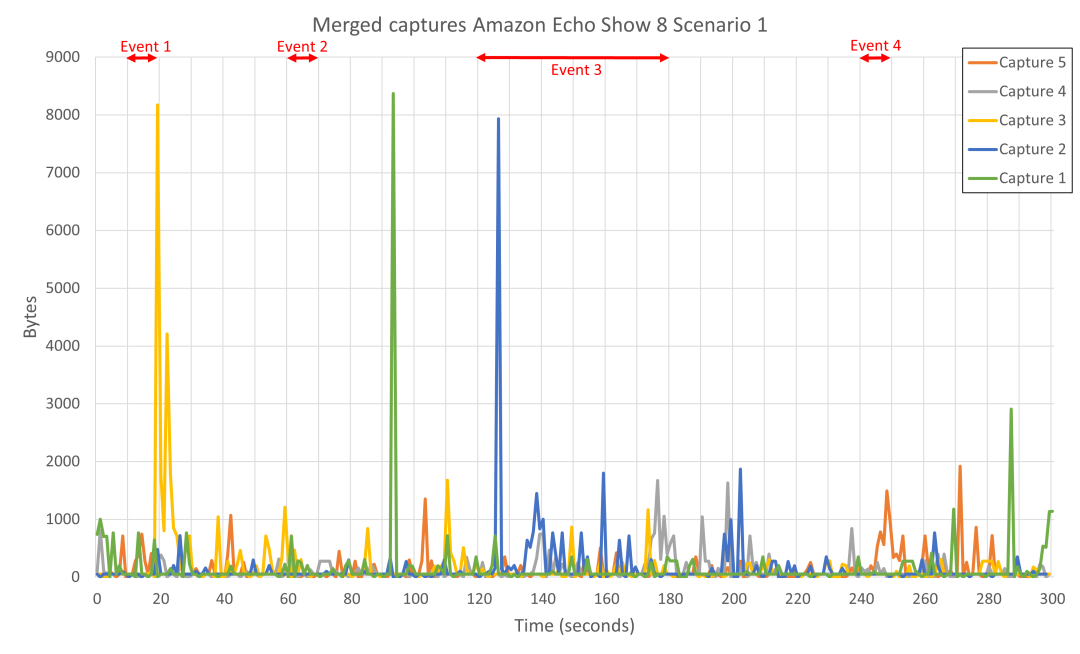


Figure 5.4: The total number of bytes transmitted per second per capture session for the Amazon Echo Show 8

For the first scenario, it is evident that the two virtual assistant devices differ a lot when it comes to outbound and inbound packets. The Amazon Echo Show 8 also display a reverse behavior with regards to null data packets when compared to the Google Nest Hub, as can be seen in Table 5.2 and Table 5.1. This could be attributable to different design requirements, and different architectural choices in the devices on how to manage communication within the Wi-Fi protocol.

5.2 Scenario 2

In Scenario 2, the Amazon Echo Show 8 and the Google Nest Hub were both set to an unmuted state as was described in Chapter 4, and the sequence of events are carried out interactively with the devices through voice commands.

A noticeable pattern is becoming pronounced together with the occurrence of events in Scenario 2. In both Figure 5.5 and Figure 5.7, spikes in packet traffic can be seen in relation to the start and the end of events for the two virtual assistant devices. The red arrows in the figures for event one, two, and four spans across ten seconds. The red arrow for the third event has a duration of 60 seconds. The duration was chosen for good measure and to better align with the grid for visual representation. There is no certain way to know for sure when an event ends, other than observing a reduction in the number of packets or bytes that seek to approach the baseline traffic pattern in between events. Even if the patterns fall slightly outside the immediate area of marked events, they can still make up a pattern.

Also keep in mind that events are initiated on a best effort basis as described in Chapter 4.

The Google Nest Hub is quite busy even outside events with sending and receiving packets as shown in Figure 5.5. Spikes in packet traffic are quite uniform. The audio call in event three is visually quite differing to the other events, settling in on around 100 packets or more per second across most captures. This gives the third event a visually distinct look from the other events and the baseline packet traffic in between events. In Table 5.3, there is a sharp reduction in the share of null data packets. This can be explained by the fact that meaningful packet exchange that contains actual data has increased due to the device now being able to record and transmit voice commands. The share of outbound packets still remain high from the previous scenario, but the share of inbound packets has increased. This could be because the device now receives more data which in turn lowers the share contributed by null data packets. The device in total send and receive 29184 packets on average across the combined captures.

The byte traffic per second for the Google Nest Hub as shown in Figure 5.6 continue to follow the packet traffic per second as shown in Figure 5.5 closely.

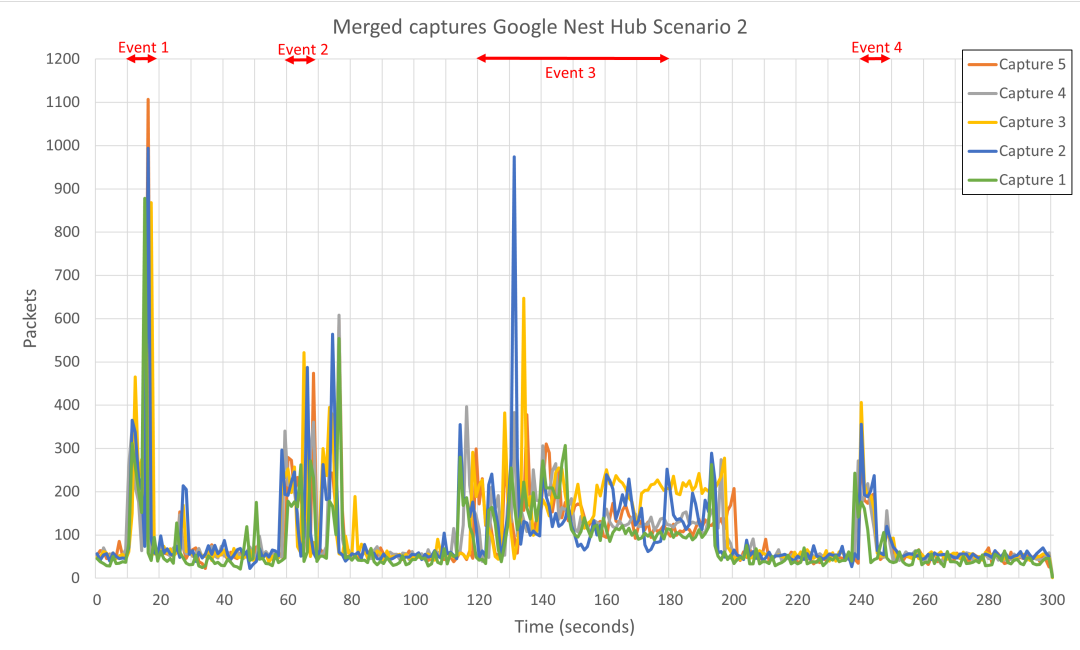


Figure 5.5: The total number of packets transmitted per second per capture session for the Google Nest Hub

Table 5.3: Statistics for the Google Nest Hub Scenario 2

Google Nest Hub Scenario 2					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	37,2 %	30,8 %	28,7 %	31,7 %	31,9 %
Non-null data bytes	94,5 %	95,9 %	95,9 %	95,4 %	95,4 %
Outbound packets	70,6 %	63,2 %	58,1 %	63,5 %	63,7 %
Outbound bytes	53,5 %	45,5 %	42,8 %	46,4 %	46,4 %
Inbound packets	29,4 %	36,8 %	41,9 %	36,5 %	36,3 %
Inbound bytes	46,5 %	53,5 %	57,2 %	53,6 %	53,6 %
Total # of packets	24334	30347	33004	29586	28647
Total # of bytes	7632346	10351711	10508143	9319723	9072048

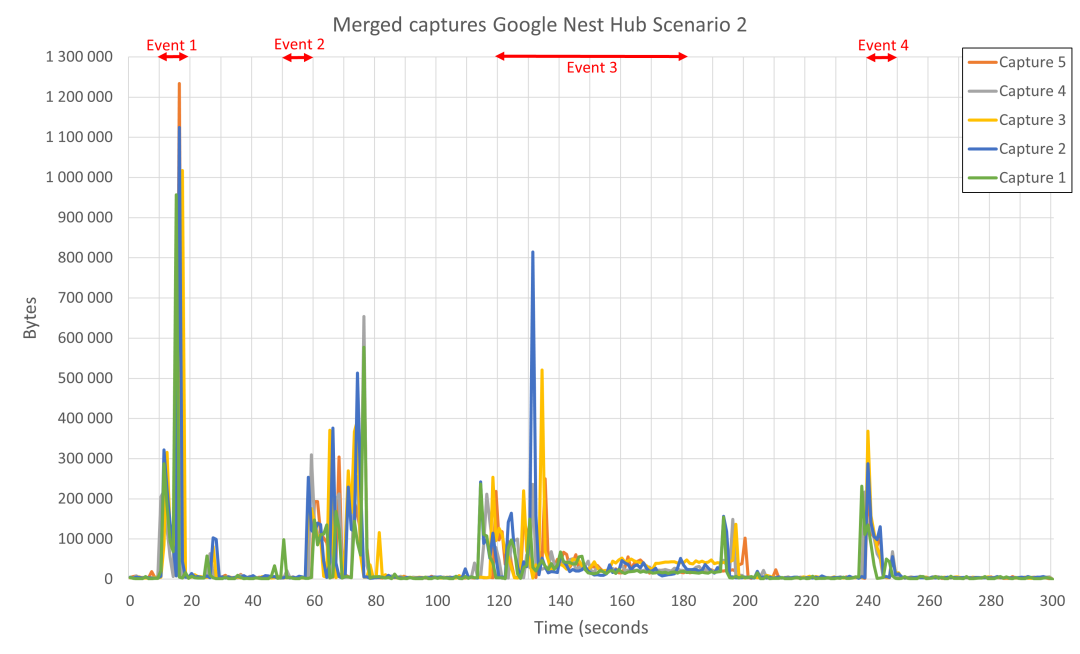


Figure 5.6: The total number of bytes transmitted per second per capture session for the Google Nest Hub

The Amazon Echo Show 8 is characterized by that the majority of packets and bytes are inbound. This was also the case in the previous scenario, only that there is now mostly a slight increase across captures. Another observation in Table 5.4 is that the share of null data packets has plummeted even further to practically non-relevant values from values that were already low for the previous scenario. In Figure 5.7 there are now patterns or segments that make themselves stand out according to each event that is commenced. The pattern of the third event is clearly distinguishable from the pattern of the remaining events. The device in total send and receive 9485 packets on average across the combined captures.

The byte traffic shown in Figure 5.8 also continue to closely follow the patterns in the packet traffic as shown in Figure 5.7.

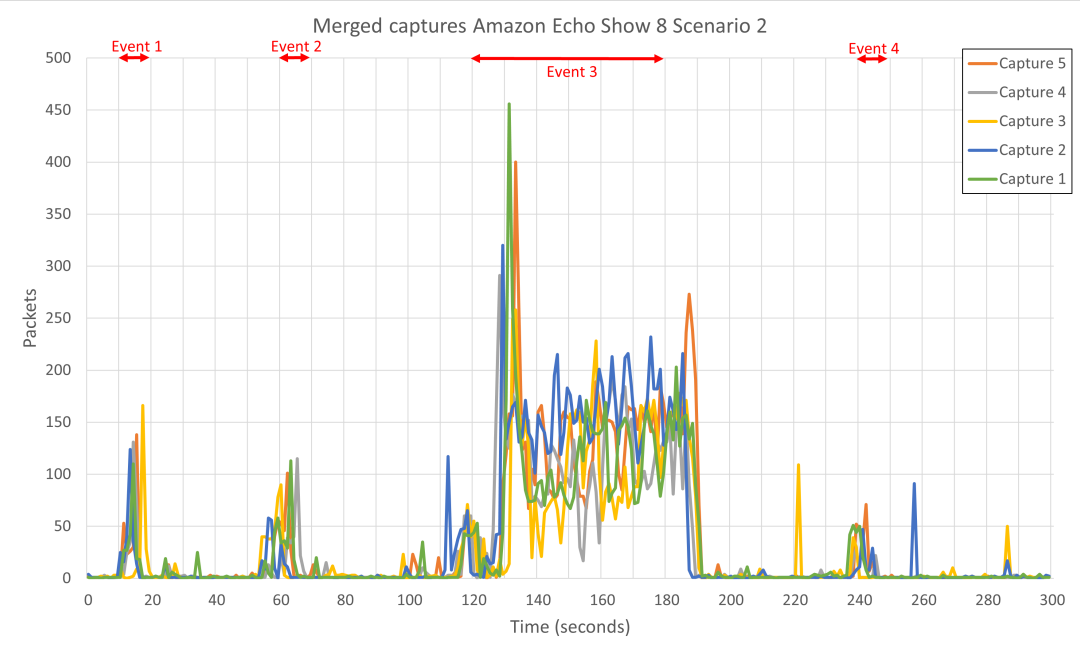


Figure 5.7: The total number of packets transmitted per second per capture session for the Amazon Echo Show 8

Table 5.4: Statistics for the Amazon Echo Show 8 Scenario 2

Amazon Echo Show 8 Scenario 2					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	0,0 %	0,1 %	0,1 %	0,1 %	0,1 %
Non-null data bytes	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %
Outbound packets	4,1 %	17,1 %	18,2 %	7,6 %	8,4 %
Outbound bytes	4,6 %	20,8 %	15,6 %	6,5 %	6,3 %
Inbound packets	95,9 %	82,9 %	81,8 %	92,4 %	91,6 %
Inbound bytes	95,4 %	79,2 %	84,4 %	93,5 %	93,7 %
Total # of packets	9155	10392	10825	8423	8630
Total # of bytes	2641097	3159989	2831295	2347150	2511405

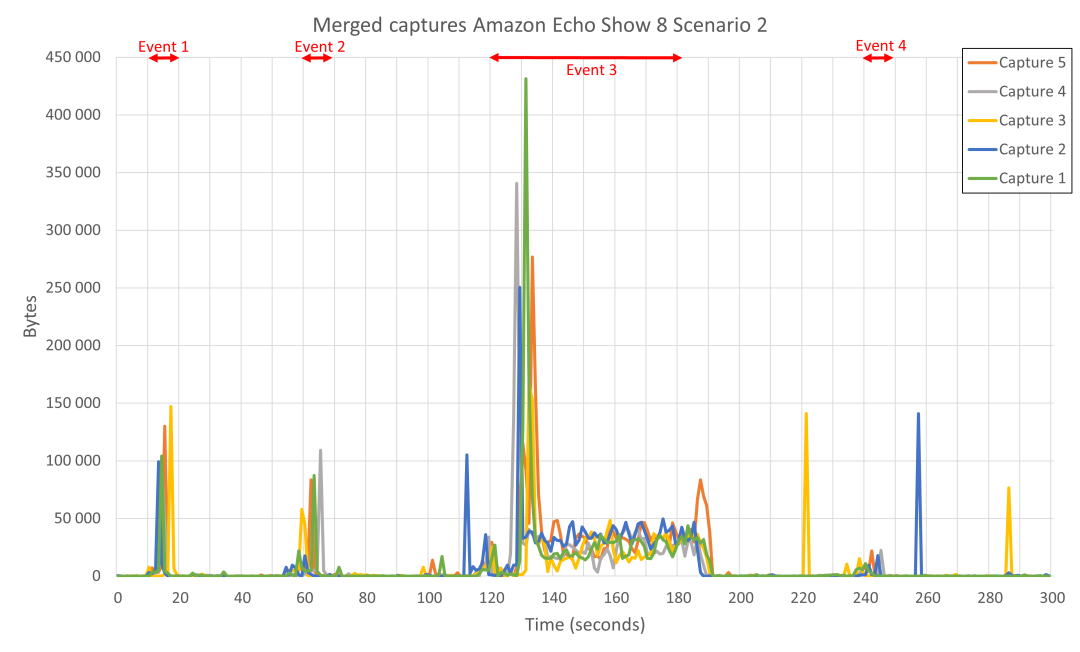


Figure 5.8: The total number of bytes transmitted per second per capture session for the Amazon Echo Show 8

It is evident that both devices have begun communicating with other entities in this scenario. The visual approach shows that patterns are forming for both devices. The Google Nest Hub device is still more talkative than its counterpart in that it sends and receive quite a lot more packets and bytes, close to three times as many.

5.3 Scenario 3

In this scenario, the Amazon Echo Show 8 and the Google Nest Hub were both set to a muted state. Furthermore, the smartphone with the associated companion apps for the Amazon and Google device was connected to 4G exclusively. In contrast to the preceding scenarios, this scenario was carried out by commencing each event in the sequence of events from the companion apps. Which means no voice commands to initiate any of the events from the physical domain was passed on to the devices in this scenario.

Patterns still coincide with the events for the Google Nest Hub as shown in Figure 5.9. However, there are some differences from Figure 5.5 in the previous scenario. The number of packets required to carry out the first event has been reduced quite drastically. The second event remains quite stable across the two figures. The third event also appear stable, in particular during the period where the audio call has stabilised after establishing the connection between the two devices. Looking at Table 5.5, the percentages stay quite similar overall to Table

5.3 in scenario 2. The device in total send and receive 23721 packets on average across the combined captures.

The byte traffic per second for the Google Nest Hub as shown in Figure 5.10 continue to follow the packet traffic per second closely as shown in Figure5.9.

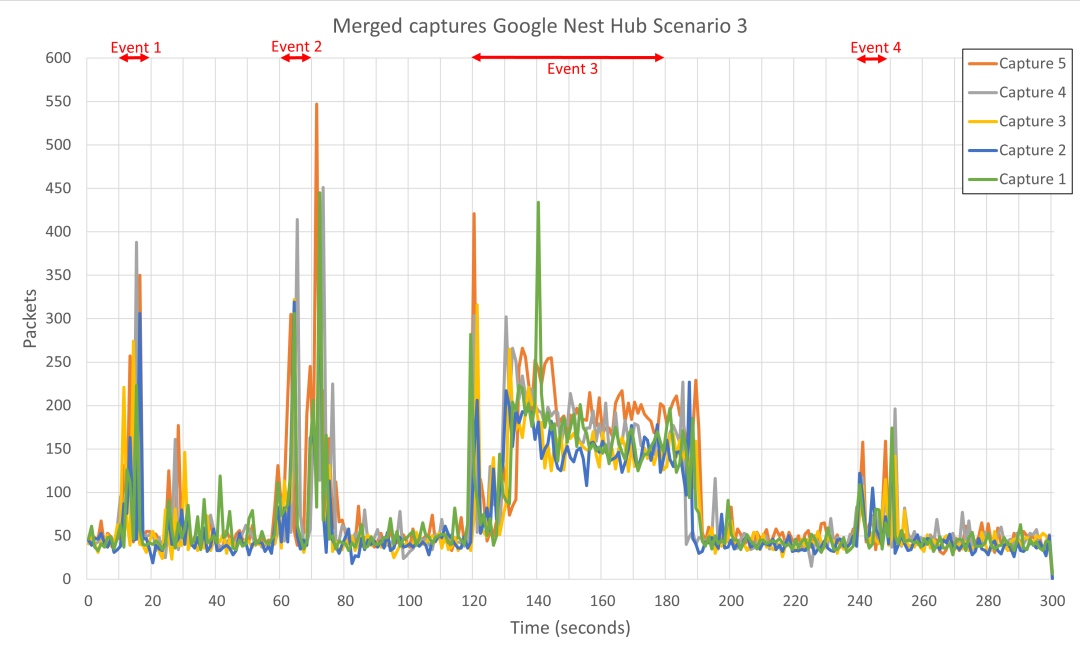


Figure 5.9: The total number of packets transmitted per second per capture session for the Google Nest Hub

Table 5.5: Statistics for the Google Nest Hub Scenario 3

Google Nest Hub Scenario 3					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	36,9 %	40,0 %	39,7 %	37,4 %	34,7 %
Non-null data bytes	91,9 %	91,0 %	90,9 %	91,9 %	93,0 %
Outbound packets	68,7 %	71,2 %	70,4 %	66,5 %	59,9 %
Outbound bytes	40,3 %	43,5 %	43,5 %	39,9 %	33,8 %
Inbound packets	31,3 %	28,8 %	29,6 %	33,5 %	40,1 %
Inbound bytes	59,7 %	56,5 %	56,5 %	60,1 %	66,2 %
Total # of packets	23564	21046	22229	24767	27001
Total # of bytes	4964152	4393775	4473126	5273544	6179222

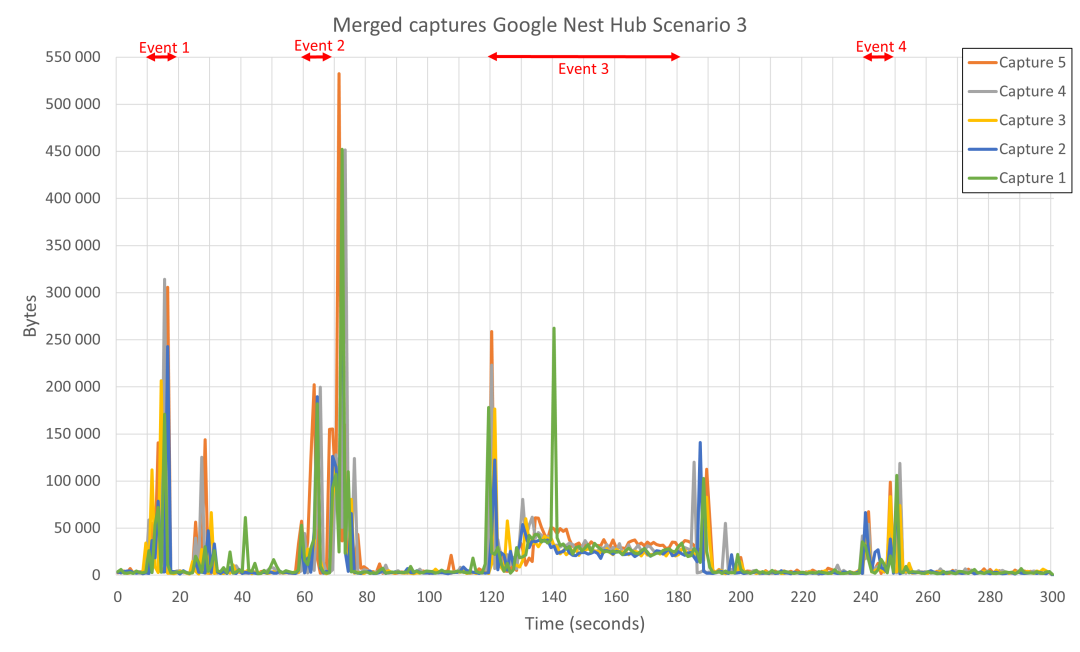


Figure 5.10: The total number of bytes transmitted per second per capture session for the Google Nest Hub

In Figure 5.11 and 5.12, there is a new development for the Amazon Echo Show 8. Compared to the Figures for the Amazon Echo Show 8 on packet and byte traffic per second in scenario 2, event patterns are now less distinct. There is an increase in seemingly random spikes of packet traffic outside the events. The packet traffic is also quite low across all the events. The device in total send and receive 2344 packets on average across all captures. However, even though the packet traffic and the byte traffic is lower in total compared to packet traffic for the device in scenario 2, Table 5.6 does not reveal any noticeable differences in the overall share of outbound or inbound packets.

Table 5.6: Statistics for the Amazon Echo Show 8 Scenario 3

Amazon Echo Show 8 Scenario 3					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	0,3 %	0,3 %	0,2 %	0,2 %	0,2 %
Non-null data bytes	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %
Outbound packets	14,5 %	16,0 %	12,5 %	28,5 %	9,6 %
Outbound bytes	27,9 %	33,0 %	23,9 %	50,1 %	20,5 %
Inbound packets	85,5 %	84,0 %	87,5 %	71,5 %	90,4 %
Inbound bytes	72,1 %	67,0 %	76,1 %	49,9 %	79,5 %
Total # of packets	2196	2267	2510	1892	2855
Total # of bytes	727572	706743	779610	543479	850960

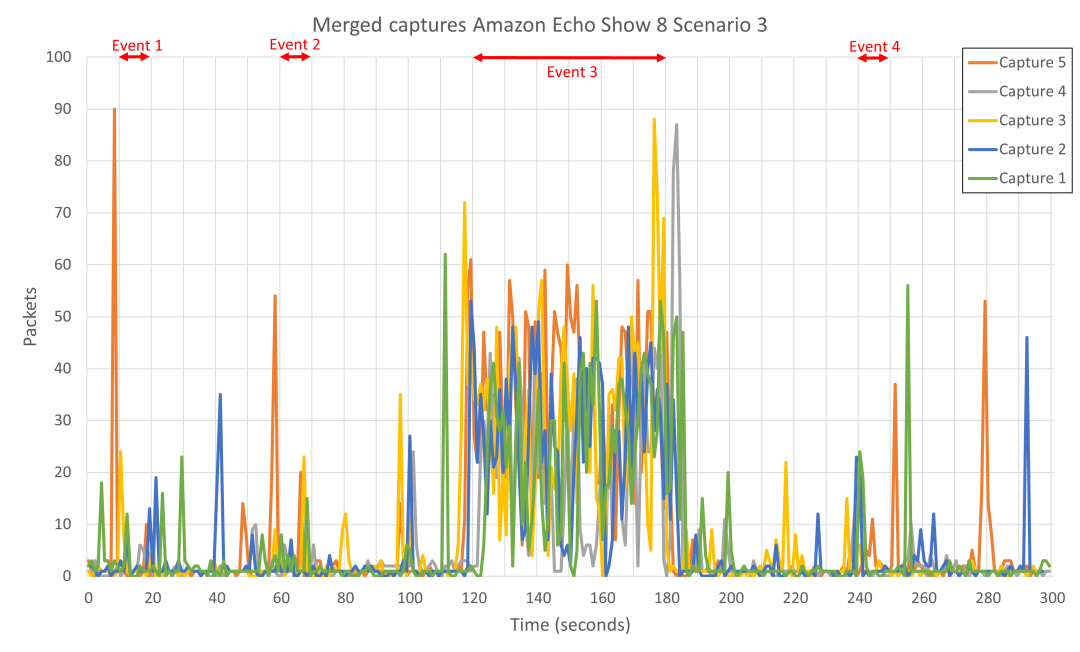


Figure 5.11: The total number of packets transmitted per second per capture session for the Amazon Echo Show 8

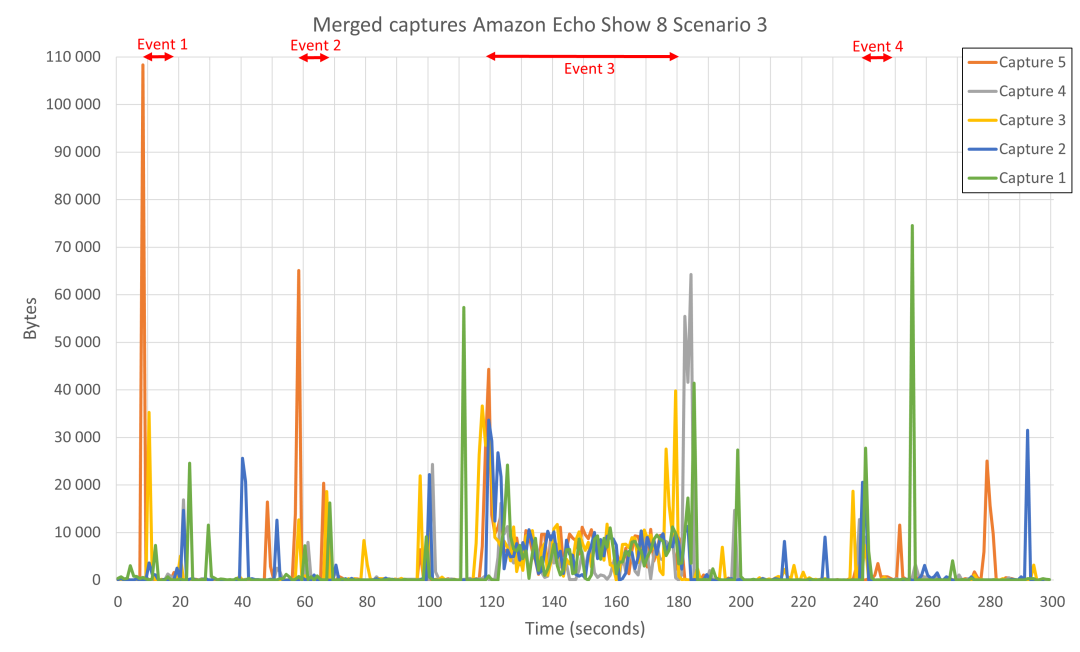


Figure 5.12: The total number of bytes transmitted per second per capture session for the Amazon Echo Show 8

The Google Nest Hub is subject to a noticeable decrease in packet traffic in

scenario 3 compared to scenario 2, but it remains high. The decrease in packet traffic for the Amazon Echo Show 8 is more drastic, amounting to only one-third of the traffic compared to the previous scenario.

5.4 Scenario 4

In the final scenario, the Amazon Echo Show 8 and the Google Nest Hub were both set to an unmuted state. In the same manner as in the previous scenario, the smartphone is connected to 4G exclusively.

The Google Nest Hub does not exhibit much change visually compared to the previous scenario. In Figure 5.13, the patterns are quite similar to what was presented in scenario 3 for this device. Looking at the Table 5.7 there is no pronounced differentiation to the previous scenario when it comes to outbound or inbound packets either. The total number of packets across captures is quite stable.

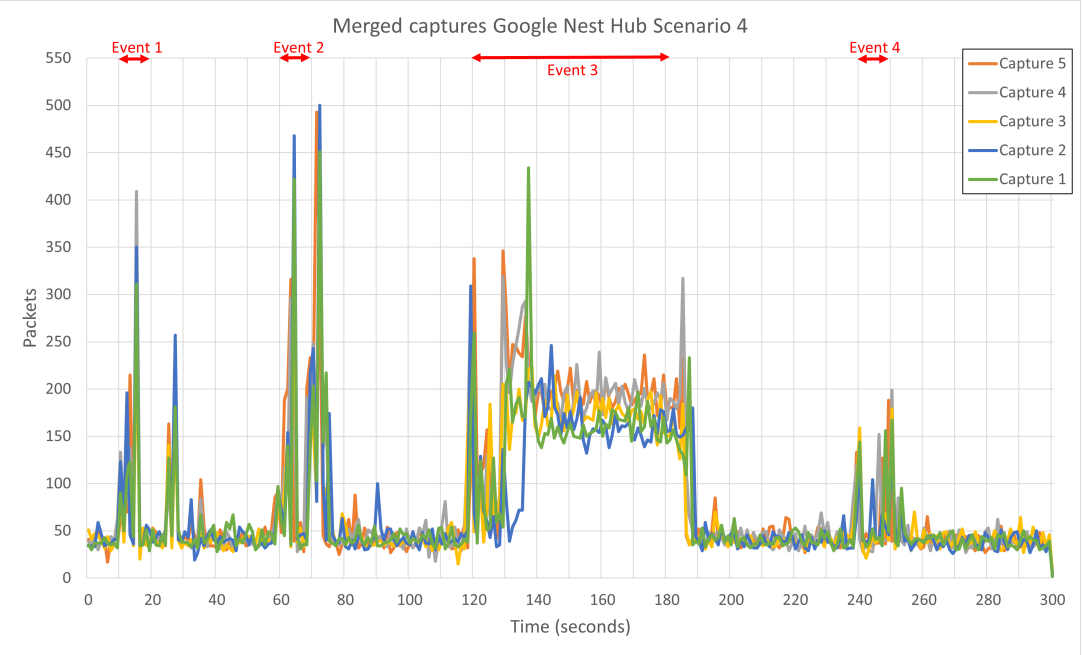


Figure 5.13: The total number of packets transmitted per second per capture session for the Google Nest Hub

Table 5.7: Statistics for the Google Nest Hub Scenario 4

Google Nest Hub Scenario 4					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	42,6 %	43,7 %	42,2 %	37,4 %	37,2 %
Non-null data bytes	91,1 %	90,8 %	90,8 %	92,7 %	92,9 %
Outbound packets	69,6 %	70,9 %	69,0 %	61,3 %	60,2 %
Outbound bytes	38,8 %	38,8 %	40,0 %	33,1 %	31,2 %
Inbound packets	30,4 %	29,1 %	31,0 %	38,7 %	39,8 %
Inbound bytes	61,2 %	61,2 %	60,0 %	66,9 %	68,8 %
Total # of packets	22875	22447	22810	25463	25528
Total # of bytes	5066723	4889340	4820990	5982059	6157768

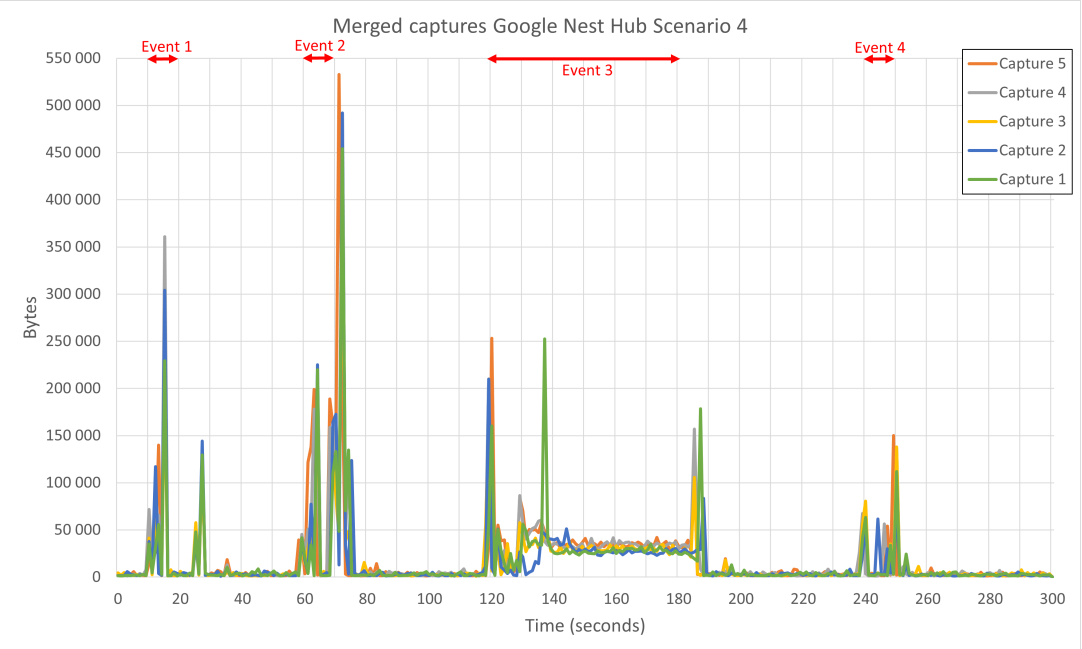


Figure 5.14: The total number of bytes transmitted per second per capture session for the Google Nest Hub

For the Amazon Echo Show 8, it can be observed in 5.15 that the first event, and third event, has increased from the corresponding figure in the previous scenario. The first has increased only slightly. The second event on the other hand, increased by a large margin. For the fourth event in the figure, it is more difficult to establish if a pattern can be observed. For the statistics, a new behavior is observed. In Table 5.6 belonging to scenario 3, the amount of outbound packets was very low. In Table 5.8 it is evident that the outbound packets now make up the majority of the packet traffic.

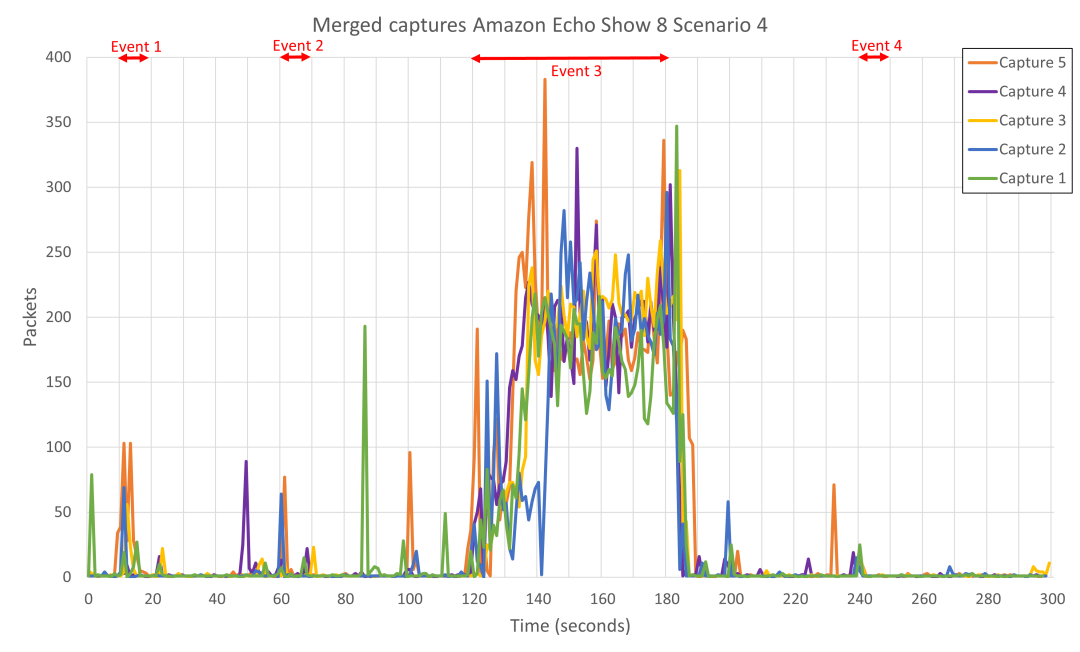


Figure 5.15: The total number of packets transmitted per second per capture session for the Amazon Echo Show 8

Table 5.8: Statistics for the Amazon Echo Show 8 Scenario 4

Amazon Echo Show 8 Scenario 4					
	Capture 1	Capture 2	Capture 3	Capture 4	Capture 5
Null data packets	0,1 %	0,1 %	0,1 %	0,1 %	0,1 %
Non-null data bytes	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %
Outbound packets	69,6 %	64,0 %	62,2 %	66,3 %	56,7 %
Outbound bytes	71,0 %	79,9 %	82,4 %	85,9 %	61,3 %
Inbound packets	30,4 %	36,0 %	37,8 %	33,7 %	43,3 %
Inbound bytes	29,0 %	20,1 %	17,6 %	14,1 %	38,7 %
Total # of packets	10137	9972	11424	11717	12614
Total # of bytes	2657342	4455183	5192247	5622733	5010338

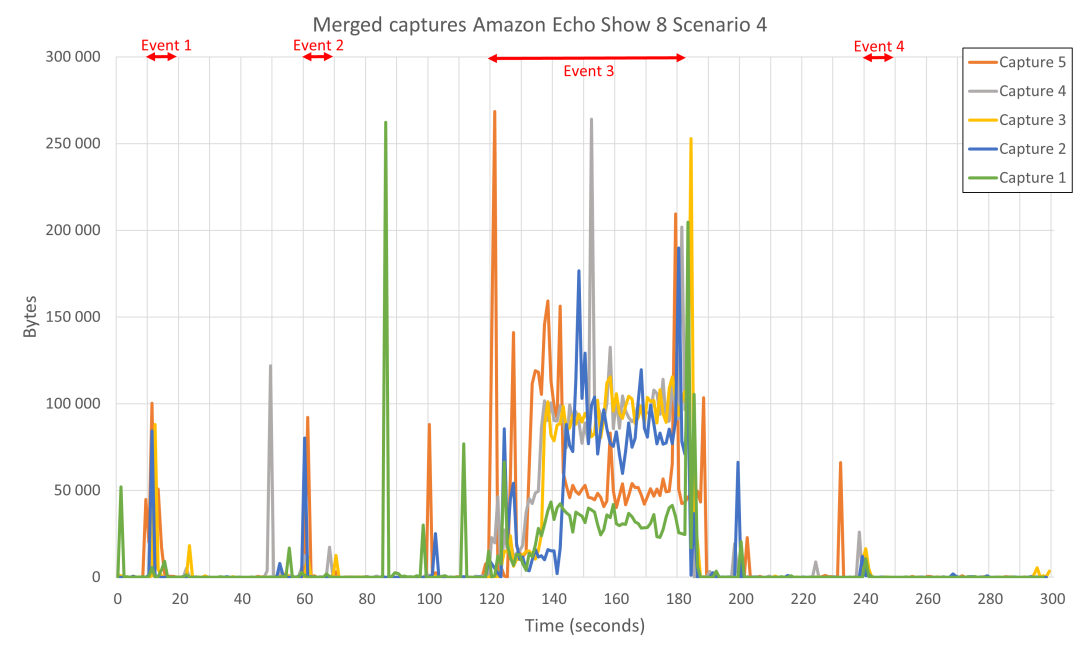


Figure 5.16: The total number of bytes transmitted per second per capture session for the Amazon Echo Show 8

In the fourth scenario, the Google Nest Hub continue to closely follow scenario 3 both graphically and statistically. The Amazon Echo Show 8 on the other hand, is heavily influenced by moving from scenario 3 both graphically and statistically.

5.5 Results

Visual graphs have been interpreted to find occurrences of patterns which can help infer ideas about the different scenarios. In addition, tables with statistics such as outbound and inbound packets, total number of packets, and total number of bytes have been presented.

The Amazon Echo Show 8 displayed large differences going from scenario 3 to scenario 4 when looking at Table 5.6 and 5.8. The number of packets and bytes in total increased. The percentages for the outbound and inbound traffic for both bytes and packets also changed, with outbound traffic increasing and inbound traffic decreasing in scenario 4. Comparing Figure 5.15 and Figure 5.15 of the fourth scenario to 5.15 and 5.15 in the third scenario, it is evident that the third event was affected by moving from scenario 3 to scenario 4 for the Amazon Echo Show 8. In scenario 4, the virtual assistant device is put an unmuted state, and the pattern for the third event increase by a large margin. Seeing as the only difference between the two scenarios was whether the virtual assistant device was muted or unmuted, it can be assumed that the Amazon Echo Show 8 do not bother with transmitting packets and accompanying bytes while it has assumed the muted state of scenario 3.

For the Amazon Echo Show 8, scenario 2 and scenario 4 can be differentiated from each other as the percentage of inbound packet traffic in scenario 2 is much greater in 5.4 than in 5.8. Which also means that the percentages of outbound packet traffic is much greater in 5.8 than in 5.4. This implies that when there is a device on a different network from the one being monitored, communicating with the Amazon Echo Show 8 and carrying out all commands from the other device, the inbound traffic is quite large for this virtual assistant device. An attacker could use this information to infer that a user affiliated with the smartphone for communication with the Amazon Echo Show 8 is likely to not be at home.

The same pattern can not be applied as conclusively to the Google Nest Hub for scenario 2 and 4, as the inbound packet percentages stay very close to each other across scenario 2 and scenario 4, and the same applies to the outbound packet percentages in both scenarios. However, it can be observed that the percentage of bytes belonging to the inbound traffic increase by around 10% in scenario 4 compared to scenario 2. In addition, the percentage of inbound packets is slightly lower on average in scenario 4 when compared to scenario 2. This could be used by an attacker to infer with a slightly higher degree of confidence that an affiliated party is connected to a network separate from the virtual assistant device, and possibly being absent from the home. The Google Nest Hub also put itself in opposition to the Amazon Echo Show 8 when moving from scenario 3 to scenario 4.

The Google Nest Hub is unique across all scenarios presented in that the percentage share of inbound packet traffic from its packet captures, never reach or surpass the percentage share observed by the outbound packet traffic. This make it so that it is difficult to use the notion of outbound and inbound packet traffic to

make informed assumptions about this virtual assistant device.

In Table 5.9, the presence of the smartphone used for the research can be seen, presented in the table by its unique Media Access Control (MAC) address.

Table 5.9: Smartphone presence across scenarios

MAC address	2e:84:11:fb:b4:31
Google Nest Hub Scenario 1	No
Google Nest Hub Scenario 2	Yes
Google Nest Hub Scenario 3	No
Google Nest Hub Scenario 4	No
Amazon Echo Show 8 Scenario 1	No
Amazon Echo Show 8 Scenario 2	Yes
Amazon Echo Show 8 Scenario 3	No
Amazon Echo Show 8 Scenario 4	No

If an adversary is to learn the MAC address of a smartphone belonging to a user of a household, the adversary can use this as an advantage. An adversary may have already eavesdropped on a target household and its associated WLAN over a period of time, and performed multiple packet captures. This adversary may in future packet captures infer whether or not particular MAC addresses are missing from the packet captures by comparing the new packet captures to old packet captures. This could assist an adversary with inferring user presence.

Chapter 6

Discussion

In this Chapter, nuances of the research will be discussed. What could have been done to improve the accuracy of the research? What could have been done differently? Any unexpected change of plans?

A passive eavesdropping approach can be effective at pointing out patterns, and discovering consistencies or inconsistencies in metadata. It can also help reveal origins and destinations of network traffic, and in which directions the corresponding flow of traffic is traveling. Finally, a passive eavesdropping approach is quick to set up and initiate, and does not necessitate large computing power to carry out. However, the obtained information from this type of inference does not necessarily yield any results. The obtained information could also be difficult to interpret efficiently.

As this research was practically only carried out manually with little automation on the part of the analysis itself, there may be inconsistencies in the observed results that software and automation would have noticed more readily. There could also be results that the researcher failed to observe due to a lack of experience, or by accident. The scenarios chosen could also have been devised differently. There may have been other approaches and scenarios that would have fit the research better as well. In addition, the test bed or setup as described in 4 is experimental. This can imply that it is difficult to apply the setup used in the research to real-world conditions with a lot more variety in data sources and network behavior.

The research carried out was focusing on Wi-Fi only and as such, mainly concerned with the data link layer, also known as layer 2 in the Open Systems Interconnection model (OSI) [13]. This put MAC addresses at the center of the research for packet capturing, which may not be the most feasible approach for the research questions that were put forward.

There were multiple times when an unexpected change of plans occurred. Mostly, a slight change in strategy was necessary every time the researcher met with the tutors of the project. Successive meetings had a tendency to gradually shift both the researcher and the tutors ideas of what the project was to revolve around. Until after some time, both parties were on the same page and had a

common vision of what the thesis should contain.

The first research question from Chapter 1 asked if user presence can be inferred from passive eavesdropping on virtual assistant device network traffic. It can be argued that the research managed to answer this research question only to a certain extent, because the Amazon Echo Show 8 was found to reveal patterns in its traffic, while the Google Nest Hub was much more difficult to interpret. The second research question functioned as a supplement to the first, in that it would be beneficial to also address the literature on privacy in virtual assistant devices to broaden the scope of the research. From Chapter 3, the challenges for virtual assistant devices are many. The challenges range from the susceptibility of virtual assistant devices to be disrupted by illegitimate communication, to poor access control mechanisms and weak authorization mechanisms for users.

6.1 Countermeasures

To mitigate and reduce the potential of passive eavesdropping through traffic analysis, countermeasures such as packet padding and transmission of dummy packets are options to consider. Packet padding involves the the addition of bogus packets or bytes to the original packets to shroud the packet lengths. Dummy packets can be transmitted to obfuscate the overall traffic pattern, and make any patterns less noticeable. However, these techniques may not be included by default as they can bring with them downsides such as increased overhead, using more bandwidth, increase transmission times, or introduce delays. It may also require more energy on the part of IoT devices that are battery and resource sensitive. They may also be viewed as not important to the overall strategy of the vendors for their virtual assistant devices if it interferes too much with usability. There is also the risk that utilizing them is counterproductive in that they assist an adversary to profile the traffic because new patterns have been introduced by the countermeasure itself.

In [14], a mechanism for packet padding is proposed. This mechanism seek to pad packet lengths to make it more difficult to carry out traffic analysis. The mechanism resulted in increased delay of packet traffic [14]. The accuracy of the machine learning techniques used to classify or profile the IoT traffic in the research was reduced by at least three quarters after applying the packet padding mechanism according to the researchers [14].

Pinheiro et al. [15] suggested to implement the use of Software Defined Networking (SDN). This is a dynamic approach that takes into consideration the resource allocation of the associated network to determine whether packet padding should be carried out when comparing the overhead produced for the purpose of privacy, to the performance of the network. As an example, this solution managed to reduce the accuracy of one of the four machine learning techniques, namely the random forest machine learning technique, by approximately 90 %.

To tackle the increase in overhead from packet padding, the research in [16] goes to show that the issue with overhead can be reduced. For example utilizing

their technique padding on-demand instead of the traditional approach of pre-padding the traffic can help reduce the overhead as all original packets does not necessarily need to be padded [16]. The researchers further combined the padding on-demand methodology with what is referred to as size-based network encoding to achieve a reduction in overhead from packet padding [16].

However, the research of [17] argues that packet padding is not necessarily good enough to obfuscate traffic as they found that they could successfully classify traffic even if packet padding was applied in their experimental setup.

Chapter 7

Conclusion

This Chapter highlights the final thoughts on the research, and touches upon possible future work.

This research conducted passive eavesdropping through packet traffic capturing on two virtual assistant devices, the Amazon Echo Show 8, and the Google Nest Hub. For interpreting captured packet traffic, graphs and statistics was extracted and interpreted to gain insight on the ability of the virtual assistant devices to uphold the privacy of users, and to minimize leakage of information to adversaries, in turn strengthening the privacy of virtual assistant device users.

The Amazon Echo Show 8 revealed several concerns related to privacy in connection with the approach and methodology of this research. One concern is that an adversary may be able to infer whether or not the virtual assistant device is in a muted or unmuted state with regard to its built-in microphone. Another concern is that an adversary may be able to infer the presence or the absence of a device on the local network by observing the change in percentage of outbound versus inbound network traffic destined for the virtual assistant device.

In contrast, the Google Nest Hub was much more difficult to interpret. It showed little sign of variability in outbound versus inbound network traffic while also revealing little information graphically as well. It could be that the architecture of the Google Nest Hub is helping the device to obfuscate its network traffic more effectively compared to the Amazon Echo Show 8.

It can be concluded that within the boundaries of the test bed used in this research, the Google Nest Hub proved to be more resilient to passive eavesdropping than the Amazon Echo Show 8.

7.1 Future work

To further build on the knowledge gained from this research, it would be interesting to apply machine learning techniques in an attempt to gather information that is perhaps not feasible to obtain with a manual approach such as in [5], [4] or [8].

It could perhaps also be interesting to apply the work of Gu et al. [2], namely the attempt at creating packet vectors that combined packet sizes and direction sequences. This technique could perhaps help to gain more insight on passive eavesdropping. Although they investigated a different protocol, and identified sequences for very simple events on devices less complex than virtual assistant devices, it could still apply to, and extend this research.

Bibliography

[1] J. S. Edu, J. M. Such and G. Suarez-Tangil, 'Smart Home Personal Assistants,' *ACM Computing Surveys*, vol. 53, no. 6, pp. 1–36, Nov. 2021, ISSN: 0360-0300. DOI: 10.1145/3412383. arXiv: 1903.05593. [Online]. Available: https://dl.acm.org/doi/10.1145/3412383.

[2] T. Gu, Z. Fang, A. Abhishek and P. Mohapatra, 'IoTSpy: Uncovering Human Privacy Leakage in IoT Networks via Mining Wireless Context,' in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2020-Augus, IEEE, Aug. 2020, pp. 1–7, ISBN: 978-1-7281-4490-0. DOI: 10.1109/PIMRC48278.2020.9217236. [Online]. Available: https://ieeexplore.ieee.org/document/9217236/.

[3] N. Apthorpe, D. Reisman and N. Feamster, 'A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic,' May 2017. arXiv: 1705.06805. [Online]. Available: https://arxiv.org/abs/1705.06805v1%20http://arxiv.org/abs/1705.06805.

[4] R. B. Jackson and T. Camp, 'Amazon Echo Security: Machine Learning to Classify Encrypted Traffic,' in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, vol. 2018-July, IEEE, Jul. 2018, pp. 1–10, ISBN: 978-1-5386-5156-8. DOI: 10.1109/ICCCN.2018.8487332. [Online]. Available: https://ieeexplore.ieee.org/document/8487332/.

[5] S. Dong, Z. Li, D. Tang, J. Chen, M. Sun and K. Zhang, 'Your Smart Home Can't Keep a Secret: Towards Automated Fingerprinting of IoT Traffic,' in *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, New York, NY, USA: ACM, Oct. 2020, pp. 47–59, ISBN: 9781450367509. DOI: 10.1145/3320269.3384732. [Online]. Available: https://dl.acm.org/doi/10.1145/3320269.3384732.

[6] M. Ford and W. Palmer, 'Alexa, are you listening to me? An analysis of Alexa voice service network traffic,' *Personal and Ubiquitous Computing*, vol. 23, no. 1, pp. 67–79, Feb. 2019, ISSN: 1617-4909. DOI: 10.1007/s00779-018-1174-x. [Online]. Available: https://link.springer.com/article/10.1007/s00779-018-1174-x%20http://link.springer.com/10.1007/s00779-018-1174-x.

[7] R. Trimananda, J. Varmarken, A. Markopoulou and B. Demsky, 'Packet-Level Signatures for Smart Home Devices,' in *Proceedings 2020 Network and Distributed System Security Symposium*, vol. 2020, Reston, VA: Internet Society, 2020, ISBN: 1-891562-61-4. DOI: 10.14722/ndss.2020.24097. [Online]. Available: https://dx.doi.org/10.14722/ndss.2020.24097%20https://www.ndss-symposium.org/wp-content/uploads/2020/02/24097.pdf.

[8] C. Wang, S. Kennedy, H. Li, K. Hudson, G. Atluri, X. Wei, W. Sun and B. Wang, 'Fingerprinting encrypted voice traffic on smart speakers with deep learning,' in *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, vol. 20, New York, NY, USA: ACM, Jul. 2020, pp. 254–265, ISBN: 9781450380065. DOI: 10.1145/3395351.3399357. arXiv: 2005.09800. [Online]. Available: https://doi.org/10.1145/3395351.3399357%20https://dl.acm.org/doi/10.1145/3395351.3399357.

[9] P. Cheng and U. Roedig, 'Personal Voice Assistant Security and Privacy—A Survey,' *Proceedings of the IEEE*, vol. 110, no. 4, pp. 476–507, Apr. 2022, ISSN: 0018-9219. DOI: 10.1109/JPROC.2022.3153167. [Online]. Available: https://ieeexplore.ieee.org/document/9733178/.

[10] T. Bolton, T. Dargahi, S. Belguith, M. S. Al-Rakhami and A. H. Sodhro, 'On the security and privacy challenges of virtual assistants,' *Sensors*, vol. 21, no. 7, p. 2312, Mar. 2021, ISSN: 14248220. DOI: 10.3390/s21072312. [Online]. Available: https://www.mdpi.com/1424-8220/21/7/2312/htm%20https://www.mdpi.com/1424-8220/21/7/2312.

[11] D. Overstreet, H. Wimmer and R. J. Haddad, 'Penetration Testing of the Amazon Echo Digital Voice Assistant Using a Denial-of-Service Attack,' in *2019 SoutheastCon*, vol. 2019-April, IEEE, Apr. 2019, pp. 1–6, ISBN: 978-1-7281-0137-8. DOI: 10.1109/SoutheastCon42311.2019.9020329. [Online]. Available: https://ieeexplore.ieee.org/document/9020329/.

[12] European Commission, 'The rise of Virtual Personal Assistants,' no. January, pp. 1–6, 2018. [Online]. Available: https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/Virtual%20personal%20assistants_v1.pdf.

[13] S. Mahmood, S. M. Mohsin and S. M. A. Akber, 'Network Security Issues of Data Link Layer: An Overview,' in *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, IEEE, Jan. 2020, pp. 1–6, ISBN: 978-1-7281-4970-7. DOI: 10.1109/iCoMET48670.2020.9073825. [Online]. Available: https://ieeexplore.ieee.org/document/9073825/.

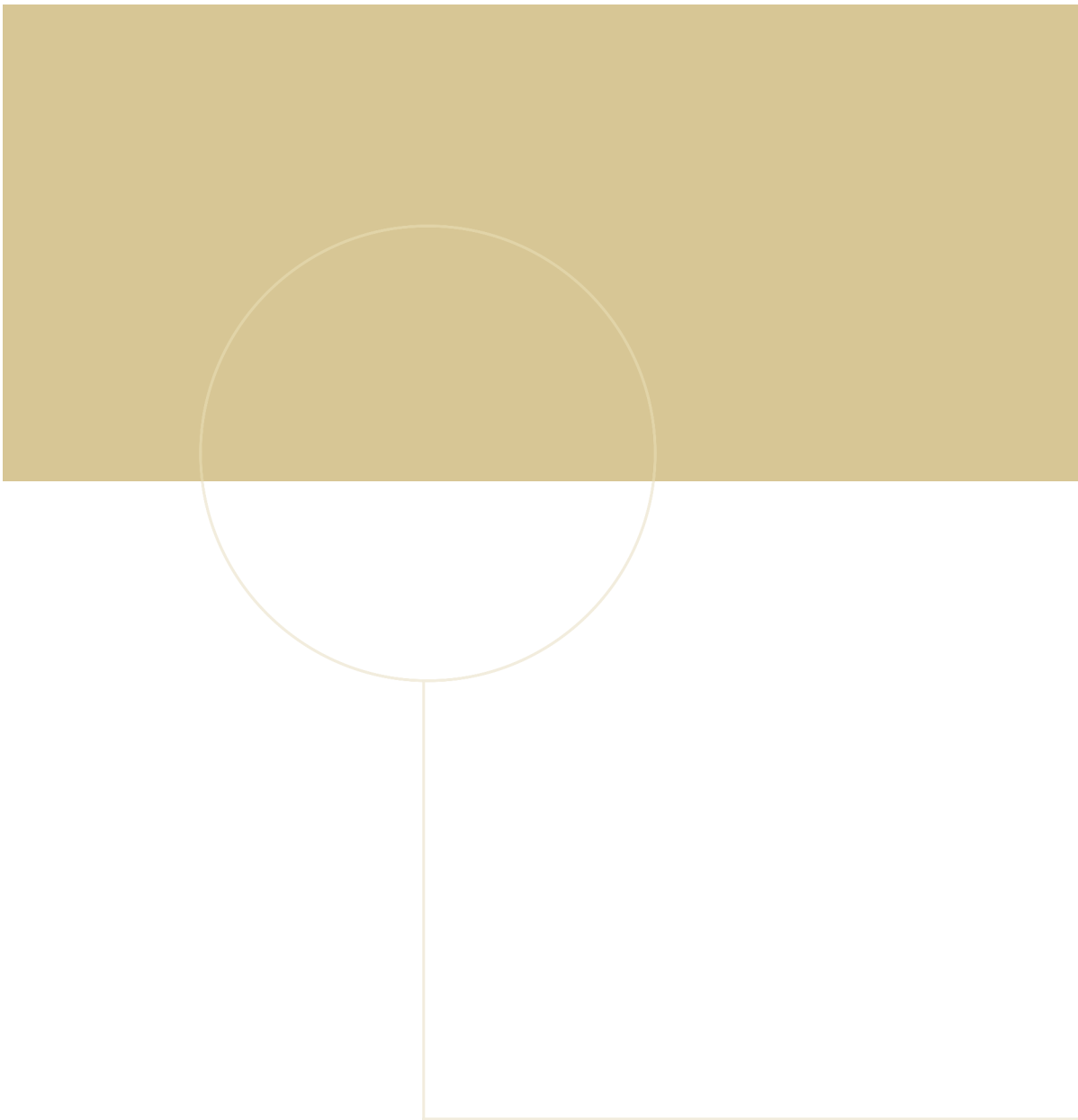
[14] A. J. Pinheiro, J. M. Bezerra and D. R. Campelo, 'Packet Padding for Improving Privacy in Consumer IoT,' in *2018 IEEE Symposium on Computers and Communications (ISCC)*, vol. 2018-June, IEEE, Jun. 2018, pp. 00 925–

00 929, ISBN: 978-1-5386-6950-1. DOI: 10.1109/ISCC.2018.8538744. [Online]. Available: https://ieeexplore.ieee.org/document/8538744/.

[15] A. J. Pinheiro, P. Freitas De Araujo-Filho, J. De M. Bezerra and D. R. Campelo, 'Adaptive Packet Padding Approach for Smart Home Networks: A Tradeoff between Privacy and Performance,' *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3930–3938, Mar. 2021, ISSN: 23274662. DOI: 10.1109/JIOT.2020.3025988. [Online]. Available: https://ieeexplore.ieee.org/document/9203848/.

[16] B. Schutz and N. Aschenbruck, 'Packet-Preserving Network Coding Schemes for Padding Overhead Reduction,' in *2019 IEEE 44th Conference on Local Computer Networks (LCN)*, vol. 2019-Octob, IEEE, Oct. 2019, pp. 447–454, ISBN: 978-1-7281-1028-8. DOI: 10.1109/LCN44214.2019.8990879. [Online]. Available: https://ieeexplore.ieee.org/document/8990879/.

[17] A. Engelberg and A. Wool, 'Classification of Encrypted IoT Traffic despite Padding and Shaping,' in *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, New York, NY, USA: ACM, Nov. 2022, pp. 1–13, ISBN: 9781450398732. DOI: 10.1145/3559613.3563191. arXiv: 2110.11188. [Online]. Available: https://dl.acm.org/doi/10.1145/3559613.3563191.



NTNU

Norwegian University of
Science and Technology