Centralised DNS-based Malware Mitigation

Examining the adoption and efficacy of centralised DNS-based malware mitigation services.

Ralph van Gurp

Technische Universiteit Delft

Centralised DNS-based Malware Mitigation

Examining the adoption and efficacy of centralised DNS-based malware mitigation services.

Master thesis submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in Complex Systems Engineering & Management
Faculty of Technology, Policy and Management

by

Ralph van Gurp

Student number: 4271742

to be defended publicly on August 2nd 2021

Graduation committee:

First supervisor, Chair	Dr. ir. C. Hernandez Gañan	Organisation & Governance
Second supervisor	Dr. ing. T. Fiebig	Information & Communication Technology
Daily supervisor, Advisor	Ir. E. Turcios Rodriguez	Organisation & Governance
External supervisor	R. Teunissen	KPN - Abuse Desk

TU Delft

Preface

First and foremost I must acknowledge the feedback, guidance, and organisational input provided by Elsa Turcios Rodriguez, whom has been involved in this work in a daily basis from its inception until its very completion, and by Raymond Teunissen at the side of KPN, whom together with others at the KPN Abuse Desk has been responsible for enabling many of the critical processes and input underlying the research. Secondly, gratitude is due to the supervisor, namely doctors Carlos Hernandez Gañan and Tobias Fiebig, for their efforts to guide and steer the research where necessary, and in their invaluable input in molding the Thesis project into a feasible and adequate work.

Furthermore, this thesis would be incomplete if I did not make mention of the circumstances under which it took place. While the Coronavirus had been rampaging for several months prior to the start of the thesis project, national measures had become increasingly strict in the months directly prior including the announcement of a national lock-down. As a result, the near entirety of the thesis project has been completed from behind desk in the city of Zaandam rather than at the office(s) of KPN. Nevertheless, I have found it interesting to receive some insight into the weekly activities at the Abuse Desk and overarching cyber security team through their daily and weekly meetings and I am grateful for this peek behind the curtain at KPN.

One of the core aspects of science is the constant improvement and expansion of knowledge by building upon the work of those that have gone before. The great Isaac Newton - to whom we owe much of our modern understanding of physics, mathematics, optics, and more - presumably uttered the following words in 1675: “If I have seen further it is by standing on the shoulders of Giants”. It is this humbleness that embodies true wisdom; the understanding that one’s achievements are not only theirs, but also those of they who have gone before, who have toiled and sacrificed to bring us to where we are now. Newton was not the first to whom such words are attributed, a 13th century French philosopher is generally recognized as the source of the ‘shoulders of Giants’ expression, but perhaps therein Newton truly embodies its essence.

If anything, the completion of this thesis is an affirmation of the strength of the shoulders upon which I stand. The shoulders of those that have gone before to carry the body of knowledge to where it is now, that have enabled the research presented in this report, and that have affected it in any which way. Perhaps more importantly, it affirms the strength of the shoulders of those who have made personal sacrifices for the opportunities that are presented to me at present; to graduate with the distinction of Master of Science at this institute. It is to them, to my family, that I owe the greatest debt of gratitude.

Ralph van Gurp
Zaandam, July 2021

Summary

Malicious cyber activity, primarily in the form of various types of malware installed on vulnerable devices connected to the internet, presents a problem with an increasingly devastating impact. Owners of compromised devices are often unaware of malware infections; a problem that is aggravated by the rise of the internet-of-things (IoT). IoT devices are often more vulnerable than traditional computing devices due to low levels of security afforded out-of-the-box, the absence of tailored security solutions, and unobtrusive operation that does not elicit frequent interaction from the user.

As a result, security measures to protect these often vulnerable devices are both more necessary and less accessible to end-users. Centralized security measures - which operate not in a distributed fashion on individual devices but at a single point to protect multiple users or devices at once - provide an opportunity for internet connected devices to be protected en masse. Internet service providers such as KPN occupy a key position in the internet hierarchy that allows them to implement such centralized measures effectively. KPN's malwarefilter presents a DNS-based implementation of this very concept, eliminating DNS requests for malicious domains, thereby limiting malware's ability to spread by frustrating compromised devices' ability to download malicious payloads or locate command and control servers.

Despite the apparent benefits, the adoption of centralized security measures remains low. This study examines the adoption of centralized security measures among KPN customers in order to gain greater insights into end-users' perception of online threats and the security of internet connected devices, their motivations for (non-)adoption of centralized security measures such as the KPN malwarefilter, and the efficacy of such services in a live environment.

The problem is investigated using a mixed-methods research design. Qualitative data is collected through research interviews, which is analysed using the Thematic analysis method. A conceptual model to assess the adoption of the malwarefilter, rooted in established theories on the adoption of security measures such as Protection Motivation Theory and the Technology Threat Avoidance Theory, is validated through regression analysis of an online survey distributed among KPN customers. Malware infection data and DNS request logs are subjected to exploratory and statistical analysis to provide further insights into the efficacy of the malwarefilter in and to characterise the internet activity and behaviour of users of the malwarefilter.

Trust in the ISP, affiliated parties, and the associated privacy implications are identified as important themes in the adoption of the malwarefilter. Reservations about the efficacy of such services, as well as the perceived implications for legitimate activities are considered by many respondents to be barriers to their adoption. Conversely, the security provided by additional security measures are often cited as reasons for their use alongside the benefits of delegating security to an apparently competent and trustworthy party, reducing the need to manage and maintain security services by the end-user themselves.

Another prevalent theme among a variety of topics is the difficulty experienced by customers in assessing the dangers posed by online threats, the security of their internet connected devices and the adequacy of their own security measures, and the assessment of the malwarefilter as an effective measure to counteract malicious activity. In line with this difficulty to assess threats and security, overestimation of one's ability to protect internet connected devices or the adequacy of existing measures may present a further barrier to the adoption of centralised security measures and can be identified as a reason for the recurrence of infections among subjects.

The perceived efficacy of the malwarefilter, subjective norms and perceptions regarding the use of security measures, and users' perceived self-efficacy were found to be significant predictors for the intention to use the malwarefilter. In line with several previous works examining the adoption of (distributed) anti-malware services, threat perceptions as well as perceptions of the perceived costs of the service are not found to be significant predictors for the intention to adopt the malwarefilter by end-users. Contrary to the identification of trust as an important theme in the analysis of the research interviews, the quantitative data provides no evidence for trust as a significant determinant for the use of centralized security measures.

An analysis of the occurrence of malware infections and their remediation time in two consecutive three-month periods, preceding and following the introduction of the intervention respectively, indicates that the malwarefilter is significantly effective in reducing end-users' vulnerability to malware threats. While the number of infections identified during the period following the intervention were notably lower for both users of the malwarefilter and non-users than for customers that were not contacted, customers who did not enable the service were found to significantly much more likely to incur a malware infection than those who did enable the service. No statistically significant effect on infection duration was found.

Exploration of the DNS log data revealed that customers that enabled the malwarefilter of primarily use their internet connection for entertainment and productivity purposes, possibly explaining the blocking of seemingly-illegitimate traffic as a potential cost associated with the service. Despite the efficacy of the malwarefilter in reducing the number of infections among users of the service, a number of suspected and known-malicious domains - which were nevertheless not tagged as malware - were identified in the examined DNS traffic. Detection methods based on temporal relations between queries and domain name composition, as proposed in some earlier works, may significantly improve the protective capability of services like the malwarefilter. Privacy-preserving technologies such as VPNs, in turn, pose a potential problem for ISPs in the provisioning of centralised DNS-based security to customers.

Contents

List of Figures

List of Tables

1

Introduction

Delineated in this chapter are the motivation and methods for research into the topic of centralized, DNS-based malware mitigation. The structure of the chapter is as follows. Section 1.1 introduces the research context; it provides a brief overview of the dangers posed by malware and the attacks it enables, and the role of the Internet of Things in the increasingly devastating real-world impact of such attacks. Section 1.2 gives an overview of previous research efforts into the mitigation of malware activity and establishes the need for greater insight into the adoption of security measures by end-users. Section 1.3 defines the research question for which the study aims to find an answer. Section 1.4 explores the academic and societal relevance of the problem. Section 1.5 lays out the contents of the subsequent chapters.

1.1. Malware

In recent years a variety of malicious cyber activities have received widespread attention and been the target of extensive campaigns by both public and private sector actors to mitigate their impact. Many of these campaigns have revolved around social engineering; phishing, scams, and other attacks that seek to exploit vulnerabilities in the manner in which end-users engage with technology and the internet rather than weaknesses within hardware or software. Nevertheless, the exploitation of device-based vulnerabilities - despite typically requiring more extensive knowledge or skills to successfully perform - presents a danger that reaches beyond an impact on the individual.

Malware, a conjecture of malicious software, comes in a variety of forms and with different purposes. Ransomware forces victims to pay a ransom or lose their data and has become increasingly popular with the growing popularity of cryptocurrencies. Adware, perhaps somewhat less detrimentally, inserts unwanted advertisements into activities performed on a compromised device. Spyware collects and extracts sensitive data, alongside Trojans which are typically used specifically to gain unauthorized access to a device. Some malwares simply try to spread through deceiving individuals to download seemingly legitimate payloads, while more intricate types actively scan the internet or local network for vulnerable devices which it can compromise and install a payload onto. Networks of compromised devices built in such a manner can subsequently be abused by malicious actors to perform powerful attacks such as Distributed Denial of Service (DDoS) attacks against otherwise secure devices and infrastructures.

DDoS attacks are widely considered to be one of the biggest concerns for cyber security professionals (Zargar et al., 2013), and the number of DDoS attacks has been rising steadily since the turn of the millennium, with a particularly sharp increase over the past five years (Cook, 2020). Quarterly reports for 2020 published by Kaspersky note a near-doubling of DDoS attacks relative to 2019 (Kupreev et al., 2020), signalling a real and growing threat to society as many critical activities and infrastructure predominantly take place in, or rely on, the internet. Recently, the importance of reliable digital infrastructure has become even more pronounced in the wake of the Coronavirus pandemic.

One of the primary growth factors of DDoS attacks has been the rising popularity of smart-home and other Internet of Things (IoT) devices (Dickson, 2019). Connecting a broad range of traditionally ‘dumb’ devices opens up avenues for innovative services such as smart healthcare applications, environment monitoring, and smart city solutions. However, IoT devices and services are, by nature, vulnerable to malicious cyber threats as they lack the protective measures afforded within enterprise parameters (Bertino et al., 2016). The weaponisation of IoT devices received widespread attention in the aftermath of the Mirai botnet attacks in 2016 and similar attacks that have occurred since (Vlajic & Zhou, 2018). As non-traditional computing devices account for a growing fraction of the market, the potential to exploit their vulnerabilities becomes an ever greater worry (and thus, a topic of interest).

1.1.1. Economic and Societal Impact

Accompanying these developments is a significant and growing economic cost (Lafrance, 2016; Cook, 2020). DDoS attacks are frequently capable of causing hours or even days of consecutive outages at major online service providers (Sachdeva et al., 2010). Lost revenue resulting from outages can approach significant sums; the average cost of a DDoS attacks on small and medium-sized businesses and larger enterprises are estimated at $120,000 and $2 million respectively (Kaspersky, 2018).

Similarly, a shared report by the NBIP (Nationale Beheersorganisatie Internet Providers) and SIDN (Stichting Internet Domeinregistratie Nederland) estimates the total damage to Dutch companies at approximately €1 billion between July 2017 and June 2018 (Boerman et al., 2018); a figure which is bound to have grown significantly since their publication. Alongside the direct economic damage resulting from attacks, organisations may be forced to divert investment into cyber security measures and remediation efforts, and incur reputation damage among clients and partners (Abhishta, 2019).

In addition to the vast economic damage resulting directly from DDoS attacks, there is the societal cost of key services being unavailable at specific moments or for extended periods of time. Sufficiently large attacks may paralyze governments (Goth, 2007), and critical services related to healthcare, law enforcement, and disaster response may equally suffer. Previous incidents targeting (public) services include attacks against major payment providers such as MasterCard, as well as the U.S., Australian, and Irish governments (and many more), stock exchanges, major internet service providers, social media websites, and leading web hosting companies (Sachdeva et al., 2010).

1.1.2. Malware Mitigation

The mitigation of malware, both its spread and the activities it enables, is often performed at one of several potential points of impact. Individual device owners may install security software that reduce the vulnerability of their devices to compromise in a more or less passive manner (such as firewalls and virus scanners). Prospective victims may acquire security services provided by specialized security companies that more actively mitigate threats (as is often the case with DDoS protection and similar services). Governments, police forces, and other security actors may furthermore engage in operations to identify and eliminate malware networks and threat actors.

Efforts to detect and mitigate malware range from scanning individual devices for signatures of known malwares to the analysis of internet traffic for patterns or traffic features that indicate illegitimate activity. The analysis of DNS traffic in particular has previously been identified as a promising avenue towards the detection of malware networks and the mitigation of their activities (Alieyan et al., 2017). Unsurprisingly, some key parties such as ISPs and other operators of network infrastructure have started to offer centralized security services to their customers, yet the adoption of such systems remains low (Antonakakis, Dagon, et al., 2010; KPN, 2020).

ISPs occupy a position in the global 'network hierarchy' that allows them to collect and analyse data required for the implementation of such security measures. Notably, the position of ISPs allows them to implement measures on behalf of individuals that lack the required skills to alleviate security issues themselves. In addition to their ability to leverage both the knowledge and data associated with their core business, ISPs have economic, utility, and reputation incentives to protect their networks and connected devices against malware and other cyber threats (M. J. Van Eeten & Bauer, 2008).

1.1.3. A Complex, Socio-Technical Problem

In order for internet infrastructure to remain reliable in a world that increasingly utilises networked services, a high level of security of individual devices or entire networks must be ensured. A key consideration in achieving this, is how the developments around cyber security and the IoT specifically shape the complex, socio-technical environment in which they live. Restrictive solutions that favour strong security over accessibility may hamper innovation and discourage the adoption of new technologies, while overly permissive solutions may (continue to) pose a liability to the integrity and reliability of interconnected networks.

A variety of actors each play a key role in the systems that together comprise the internet and the way it is used to the benefit of society and the individual. Each of these actors has potentially differing interest, skill-sets, and tools at their disposal to shape this environment. Solutions that are disproportionately costly vis-a-vis the interests of the party that is required to implement them are unlikely to become reality. Solutions that require action beyond the comprehension or available skill-set of the involved actors are equally unlikely to see the light of day; a particular concern in high-tech domains such as cyber security where large knowledge gaps are present between the end-users and other parties in the system such as network operators, device designers, and attackers.

Thus, the solution space is limited not only by the available (scientific) knowledge and physical boundaries imposed by the system, but also by the institutional environment and actor network involved, and therefore presents a challenge to anyone that intends to contribute to the resolution of problems in the domain. In order to present a solution that is feasible, viable, and effective, one must analyze not only the problem at hand but also the environment in which it will operate.

1.2. Problem Statement

Over the years, a variety of tools have been developed by attackers seeking to execute DDoS attacks, and a plethora of defense mechanisms have been put forward that aim to exploit attack or attacker characteristics in order to mitigate or deny attacks (Zaroo, 2002; Zargar et al., 2013). Previous research has investigated technical solutions ranging from simple white and blacklisting of known-malicious sources to more complicated methods.

The implementation of centralized security measures at the site of internet service providers offer opportunities for efficient and scalable identification and mitigation of malware threats (M. J. Van Eeten & Bauer, 2008; Asghari et al., 2015). The domain name system offers opportunities for the implementation of such services; DNS is widely used by attackers to control compromised hosts and engage attacks, thus making it a suitable target for the detection of compromised devices (W. Li et al., 2019).

Earlier research on the role and abilities of internet service providers in mitigating the effects of malware infections has examined remediation-based methods (Altena, 2018; Verstegen, 2019; Bouwmeester, 2020), with the intention of improving the rate at which end-users are capable of resolving malware infections of their devices. Research on end-user security behaviour has focused primarily on the enterprise context (Y. Li & Siponen, 2011). Although the previous years have seen some studies into security behaviour and the adoption of distributed information security tools and software in a household context, little is known about consumers' concerns and motivations in the adoption of centralized security measures.

Similarly, earlier studies on DNS-based malware mitigation have taken place largely in theoretical contexts, often examining little more than the degree to which proposed algorithms or solutions are capable of detecting a number of malicious domains from among a dataset of captured DNS traffic. Assessment of the value of such measures to the end-user, in combination with an assessment of their efficacy in practice, has seen little to no study. Thus, this warrants an investigation into the adoption of centralised DNS-based security measures by end-users, and the efficacy of such measures in a live environment.

1.3. Research Question

One can easily recognize the necessity for effective measures to identify malicious activity originating from IoT devices, and the subsequent need to act on this information. The unique characteristics of IoT devices call for a more proactive approach in limiting the spread of malicious software within networks, as owners of IoT devices will often be unaware of a compromised device. Even in cases where malicious activity is monitored and end-users are actively warned about device compromise they may not have received or read the warning, have paid no attention to it, or not been notified of specific infections at all. Early detection methods and preventative measures limit the propagation of malware within a network and help identify compromised devices. Previous research indicates that DNS traffic provides a useful distinguishing feature for legitimate and illegitimate traffic; emphasizing the advantages of DNS-based remediation and mitigation platforms in providing early warnings to affected users and filtering malicious traffic. In accordance with these findings, the main research question is formulated as follows:

"What are the main concerns in using centralised DNS-based malware mitigation services, how effective are such services at reducing malicious activity, and how does the DNS activity of legitimate users compare to that of compromised devices?"

Sub Questions

In order to answer the main research question, a number of sub-questions must be answered. Firstly, it is important to understand end-users' perceptions of online threats to their internet connected devices, and whether and why they perceive such devices to be well-secured or vulnerable in general, and how this translates to their own devices. This requires an understanding of the perceived security issues with internet connected devices and the dangers posed by online threats or the degree to which users are aware of such issues at all. Additionally, the manner in which end-users experience responsibility with regards to using and securing their internet connected devices may affect security perceptions and willingness to adopt security measures.

Secondly, motivations for - and barriers to - the adoption of centralised DNS-based malware mitigation services by home-users of internet connected devices must be explored. Identified concerns and motivations can be used to assess if and why such security measures may or may not see widespread adoption among the customer base of internet providers and other parties, and how these concerns may be tackled by the providers of centralised security services.

Thirdly, the efficacy of centralised DNS-based malware mitigation measures must be evaluated in practice. The efficacy of cyber security solutions hinges on a combination of factors and developments that are difficult to reproduce in controlled environments such as labs or through the mere evaluation of an algorithm on a pre-existing dataset. The behaviour of individuals and organisational actors, as well as developments within the wider information technology and security environments may strongly affect the practical value of any proposed solution.

Lastly, the data captured or otherwise generated by centralised malware mitigation services must be explored in greater detail in order to appraise its value as a source of data for the further and earlier identification of malicious activity among a network. Malicious internet traffic identified by the service may offer opportunities to detect infected devices at an early stage by characterising differences in temporal activity patterns and DNS traffic characteristics of known-compromised devices and others.

- **SQ1:** How do end-users perceive online threats and the security of internet connected devices?
- **SQ2:** What are end-users' motivations and barriers in the adoption of centralized DNS-based security?
- **SQ3:** How effective are centralised DNS-based malware mitigation services in practice?
- **SQ4:** What differentiates the temporal activity and query characteristics of legitimate and illegitimate DNS traffic?

1.4. Relevance

The following sections lay out the relevance of the research towards advancing the current understanding of the problem. The academic relevance of the research is explored, alongside its relevance to society.

1.4.1. Academic Relevance

The research effort contributes to the advancement of the academic body of knowledge primarily through the adaption of existing theories on (cyber) security behaviour and the application of these theories to examine end-user adoption of centralised cyber security measures implemented at the site of internet service providers. Previous research efforts have focused on the adoption of security measures in enterprise environments (Y. Li & Siponen, 2011), as well as commonplace and relatively security habits in household contexts such as the use of strong passwords, firewalls, and security software (Kumar et al., 2008; Claar & Johnson, 2012). Little is known about the motivations for adoption of security measures by end-users where the implementation and management of the measures rests not with the end-user, but with third parties such as ISPs or other network infrastructure providers, despite its apparent benefits (Kritzinger & Von Solms, 2013).

Secondarily, the majority of studies on household cyber security behaviour applies to commonplace security measures which require little knowledge of their functioning to assess their impact both in terms of efficacy and impact on values such as privacy. In contrast, this study examines specifically the use of less commonplace DNS-based mitigation services, knowledge of which may be expected to lie outside of the knowledge base of the average end-user and thus present an additional hurdle to their adoption. The study examines the effects of this increased 'knowledge gap' on the motivations for adopting (or refraining from adopting) such security services.

Thus, the study aids the related academic fields by improving the understanding of end-user motivation in the adoption of cyber security measures as a whole. It also improves the understanding of the efficacy and adoption of centralised DNS-based mitigation measures in particular, providing insight into the opportunities (or lack thereof) in further research into the application of such measures by network operators.

1.4.2. Societal Relevance

The societal relevance of the problem is displayed most prominently in the fact that each of the actors, as well as society as a whole, stands to gain from widespread adoption of effective cyber security measures (with the obvious exception of those that manage and exploit botnets, and those that buy attack services). First and foremost, it yields insights which may be used to improve the adoption of security measures by end-users, and therefore improve the security of home networks.

Improved security of home networks yields two significant advantages for broader society and individual citizens: it benefits the security of devices connected to the network as a whole (opportunities to exploit other connected devices through the compromised device are cut short), and it benefits the potential victims of attacks - those that might not be vulnerable to malware infections themselves but can be harmed through other attack vectors such as DDoS - by reducing the number of vulnerable and compromised devices which can be leveraged in an attack. On a (trans)national level, one might even consider the enhanced reliability of internet-based services as a necessity to thwart malicious activity of malicious organisations and malicious state actors that pose a danger to critical and everyday infrastructures.

Internet service providers and other network operators benefit from the research through the greater insights into the efficacy of mitigation platforms (and thus; whether it makes sense to develop these services further from a technical and business perspective). Another important benefit for these organizations is that the identification of concerns amongst their customers in the adoption of these services in their current and future forms can aid in the design of more attractive (from the end-user perspective) services and to provide overall more fitting communications to their customers, as they will be better able to address customers' concerns.

1.5. Thesis Organisation

The subsequent chapters present the conducted research and its key contributions. Chapter 2 presents a literature review that examines in detail: (1) the characteristics and mechanisms of IoT malware, (2) of DNS-based protection mechanisms, and (3) user adoption of cyber security and technology innovations. Chapter 3 provides a brief explanation of the environment in which the research takes place; the system and its actors and the intervention and its effect on the system. Chapter 4 provides a detailed overview of the data collection and analysis methods employed and the sample sizes required for sufficient reliability. Chapter 5 outlines the results of the experiment, analysed using the methods described in the previous Chapter. Chapter 6 presents the synthesis of the research effort: a discussion of the synthesised findings and their implications and avenues for future research efforts.

2 Literature Review

This chapter presents the state-of-the-art in the three areas the research concerns. First, it delineates the characteristics and operations of malware and the evolution of Internet-of-Things malware (Section 2.1). Thereafter, it explores how the Domain Name System is used to mitigate the effects of malware infections and other malicious activity (Section 2.2). Finally, the knowledge base on user adoption of technology and (cyber) security services is examined in an effort to establish the leading theories and frameworks that identify key antecedents for user behaviour in information security (Section 2.3). Based on these findings a conceptual framework is constructed that captures expected predictors for the adoption of security measures, which is defined in chapter (Section 4).

The literature search is conducted based on the methods proposed by Wee & Banister (2016), who write on the general conduction of literature reviews, and Vom Brocke et al. (2015), who write specifically on conducting a literature review in the context of information systems research.

Wee & Banister propose the explicit reporting of databases used to conduct the search, of languages and keywords, and a search strategy. Two commonly used search strategies are forward snowballing, which aims to find articles that cite the paper in question, and backward snowballing, which aims to find the articles upon which the paper in question builds its argument. Search results that yield too many papers can be narrowed through the inclusion of a selection process which should be made explicit and might relate to factors such as the impact of an article (often measured through citations), the geographical area, its the year of publication, or according other criteria.

Vom Brocke et al. note the challenges posed to information systems (IS) researchers due to the often necessary crossing of disciplinary boundaries, the strong presence of trends and buzzwords, and the rapidly growing number and length of publications. Given these challenges, Vom Brocke et al. propose a series of steps that must be conducted prior, during, and after the literature search. Prior to the search, one should develop an understanding of the topic, a justification of the review, define an appropriate scope, and assess the feasibility and coverage of the search. During the search, one should test alternative approaches, justify search techniques and parameters, and define appropriate criteria for the inclusion (and exclusion) of articles. After the literature search, the accuracy of the search should be assessed and the search rigorously documented.

Section 2.1 gives a brief insight into the mechanisms of (D)DoS attacks, the role played by botnets, the danger posed by the IoT, and an overview of some prominent malware families. Sections 2.2 and 2.3 present literature reviews of DNS-based mitigation measures and user adoption of security measures respectively. In accordance with the guidelines defined by Wee & Banister (2016) and Vom Brocke et al., a literature search has been conducted using keyword search to find appropriate search strings, and trial-and-error searching for alternative phrases and search strings. The backward snowballing technique was used to further identify influential papers and theories; those articles that form the basis for the arguments used by the authors. An overview of the literature search and the literature consulted during the review can be found in Appendix A.

2.1. Malware

Malware comes in a variety of forms and with a variety of purposes, from directly harming the victim of the malware infection to incorporating a compromised devices into a network used to conduct attacks without necessarily targeting the owner of the compromised device. Kara (2019) distinguishes several basic types of malware such as spyware, adware, viruses, bots, rootkits, Trojans, worms, and crypto-malware.

While many types of malware pose a direct threat to the owner or user of a device, the perhaps most devastating malwares are those that merely leverage infected devices for future attacks. The often massive size of networks of compromised devices that participate in such attacks means that they pose an exceedingly large threat to large organisations, governments, and (critical) infrastructures that rely on internet-enabled functionality that could potentially be paralyzed by large-scale, malware-enabled attacks. Denial of Service attacks are the prime example of this, having previously been responsible for some of the most devastating cyber attacks.

2.1.1. (D)DoS Attacks

The fundamental principle of Denial of Service attacks is straightforward; one can damage a party by disabling their ability to service legitimate clients. Most commonly, these attacks exploit the limitations of computational resources required to handle seemingly legitimate requests (commonly referred to as ‘volume attacks’), or one or more of the protocols that form the basis for information exchange on the internet (so called ‘protocol attacks’). Volume attacks straightforwardly work by sending or requesting more data than the target can handle to process, thereby denying service to legitimate clients who will either have to wait for the target to finish handling all the illegitimate requests or may be unable to contact it at all. Protocol attacks are usually slightly more sophisticated, instead capitalizing on specific weaknesses in one or more of the underlying protocols of the internet such as the Internet Control Message Protocol (ICMP), the Transport Control Protocol (TCP), or User Datagram Protocol (UDP). The most popular protocol attacks exploit the three-way handshake percent in the TCP protocol to paralyze the target, or take a hybrid approach based on the exploitation of multiple protocols.

Over the years, a variety of tools have been developed by attackers seeking to execute DDoS attacks, and a plethora of defense mechanisms have been put forward that aim to exploit attack or attacker characteristics in order to mitigate or deny attacks (Zaroo, 2002; Zargar et al., 2013). Early research focused on methods such as statistical filtering/clustering methods based on packet attributes and distributions (Feinstein et al., 2003; K. Lee et al., 2008), path tracing (Yaar et al., 2003; Law et al., 2002), and resource allocation (Lau et al., 2000). However, as attacks become progressively more ‘distributed’ due to a growing number of devices and access to the internet, such methods are proving to be less effective. The recent rise of IoT and the seemingly limitless application domain of these devices only aggravates these issues.

2.1.2. The Advent of Botnets

Despite their straightforward nature, Denial of Service attacks have proven devastating. The success of these attacks is largely the result of attackers’ ability to control vast, seemingly unrelated networks of devices with few common denominators which are therefore difficult to identify and protect against. These networks of compromised devices are commonly known as ‘botnets’ and regularly consist of hundreds of thousands of devices; several historically active botnets have consisted of up to millions of devices. These devices can be of varying types and from disparate geographical locations; now potentially including a variety of device types that have not traditionally been targets for malware infections such as smart-city infrastructure.

Botnets can function based on centralised control mechanisms that distribute orders to the bots in the network, or by employing peer-to-peer mechanisms. Figure 2.1 presents a visualization of a botnet managed by a centralised control structure, where one or multiple servers disseminate infection and attack commands in a hierarchical fashion. Figure 2.2 presents a visualization of a botnet employing a peer-to-peer mechanism that allows for decentralised dissemination of commands by utilizing (a part of) the compromised hosts as control servers. centralised control structures for botnets employ hard- and software specifically tasked with managing the lines of communication between the controller and the infected hosts; these servers are generally known as Command and Control (C&C or C2) servers. These C&C servers use a variety of protocols to coordinate the bots in the network primarily; the Internet Relay Chat (IRC) protocol, the Hypertext Transfer Protocol (HTTP), or the Domain Name System (DNS).

C&C servers may also partake in the direct propagation of the malware by transferring the files necessary to complete in infection to vulnerable devices identified by the server itself, or by other bots in the network. Most highly-advanced botnets employ methods that enable self-propagation either with or without the need to communicate with the C&C server in order to compromise and infect a device. Self-propagation mechanisms tend to be relatively simple, usually relying on scanning for vulnerabilities in other devices connected to the network. Examples of such vulnerabilities are exposed ports that have been left unprotected due to configuration flaws or weak credentials such as the default username and password set by the manufacturer. Once a connection has been established with a vulnerable device, the malware is transferred either by the original bot, or an infect-command is sent to the C&C server.

2.1.3. Malware Across Environments

The growing popularity of non-traditional computation devices is one of the main drivers behind malware-based attacks becoming more powerful and malware itself becoming more difficult to contain as a direct result of the increasing size and degree of distribution of botnets. The last two decades have seen a sharp increase in both the number of different malwares that have been sampled, as well as a significant evolution in their complexity and functionality. Generational advancements of malware can be distinguished along these lines, as malware becomes decreasingly reliant on human activity in order to spread instead employing self-replication mechanisms by propagating through media files and over the internet, and ultimately developed into systems that could be

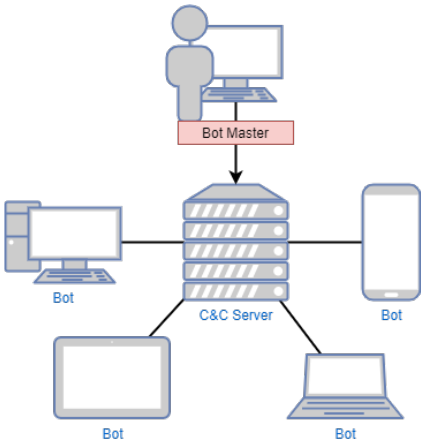


Figure 2.1: Visualization of a centralised botnet.

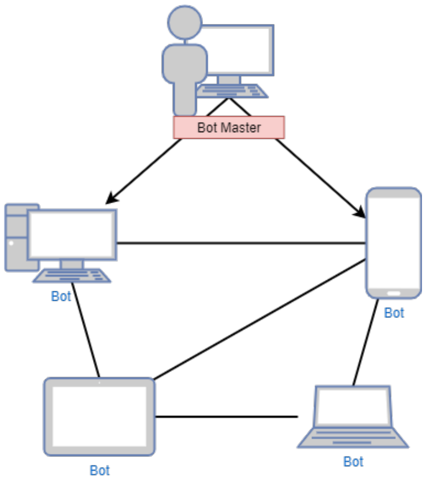


Figure 2.2: Visualization of a peer-to-peer botnet.

employed for cyberwarfare purposes and in malware-as-a-service schemes (Ligh et al., 2011).

While in the previous decades the majority of malicious software targeted Microsoft Windows (then the most popular operating system on the market), the increasingly diverse landscape of devices over recent years has shifted focus onto a variety of different operating systems and architectures. The growing use of Unix-like operating systems in smartphones and other resource-constrained devices - such as those that constitute the internet of things - has spurred the development of new malware that specifically targets these systems (Pieterse & Olivier, 2012; Ngo et al., 2020).

Mobile Malware

Zhou & Jiang (2012) present a systematic characterisation of Android malware based on installation vectors, activation method, and malicious payload. Installation vectors for mobile devices typically rely on social-engineering methods; repackaging, update attacks, and drive-by downloads. Repackaging malicious payloads into popular applications which are then distributed through either the official or an unofficial Android marketplace allows attackers to piggy-back on these popular apps and entice users to install the compromised package. Update attacks employ similar mechanisms but instead of packaging the entire malicious payload with the installation package, parts of the payload are distributed through updates to the compromised application allowing the method to circumvent traditional static scanning applications. Drive-by downloads entice users to download supposedly ‘feature rich’ apps which, for example, proceed to steal sensitive information. A number of other techniques are identified by the researchers that do not fit within any of these three social engineering approaches, for example because they openly include malicious functionality or because they require explicit installation by an actor that is aware of the malicious functionality (e.g. spyware).

Malicious payloads are typically one of four generalised types; privilege escalation, remote control, financial charge, or information collection. The analysis performed by Zhou & Jiang provides some insights into the relative prevalence of these payload types among the extensive sample. Approximately 36% of the examined malware samples incorporate at least one root exploit with the intention of privilege escalation. Remote control payloads, which turn the device into bots for remote control, were identified among 93% of the examined malware samples. Financial charge, usually performed by subscribing the user or device to attacker-controlled premium-rate services, was found among 4.4% of the examined samples. Active harvesting of various types of information such as SMS messages, phone numbers, and user accounts, was performed by approximately 60% of Android malware.

Analyses of smartphone antivirus software reveals that common solutions are only capable of detecting 20% to 80% of malware (Zhou & Jiang, 2012). Simultaneously, solutions such as dynamic code analysis which could identify active threats are typically not feasible on these devices due to limitations imposed by power consumption and computational power restrictions (Suarez-Tangil et al., 2013). The internet of things in particular presents a challenge to malware mitigation, as IoT devices are increasingly popular not only with the general public but also as potential targets for new malware.

IoT Malware

The great variety of hardware architectures employed by IoT devices, in combination with resource constraints, has spawned a large number of new malware families (Ngo et al., 2020). Kolias et al. (2017) note five reasons why IoT devices are particularly advantageous for creating botnets:

1. Constant and unobtrusive operation: IoT devices are typically always-on devices in contrast with desktops and laptops which have frequent on/off cycles.

2. Feeble protection: device vendors typically neglect security in favor of user-friendliness, stability, and time-to-market.
3. Poor maintenance: IoT devices are usually setup-and-forget and not subject to scrutiny by their owners until they stop functioning correctly.
4. Considerable attack power: IoT devices are powerful enough to produce attack traffic comparable to that of modern desktop computers.
5. Minimally interactive: infections are likely to go unnoticed since IoT devices typically require little intervention by users and often a compromise cannot be addressed short of replacing the device.

Commonplace vulnerabilities in IoT devices enable botnet masters to abuse these highly attractive targets. Bertino & Islam (2017) note that the significant heterogeneity of IoT devices complicates the use of universal security measures and update mechanisms. Physical security of such devices may be inadequate (especially in the case of devices installed in the public domain), allowing direct access to the embedded hardware and software. Additionally, these devices often simply do not allow users sufficient control and configurability to adequately enhance their security. Consequently, Bertino & Islam identify the need for intrusion-detection systems at a network level to adequately address the threat posed by malware in the IoT environment.

The majority of malicious software targeting IoT devices commonly found today are evolutions of the same source code, which is typically copied and modified to fit the needs of the attacker (Allix et al., 2014). As such, many of these malware families display a high level of similarity Ngo et al. (2020). Most IoT malware operates by scanning open internet ports commonly associated with IoT devices in an attempt to gain access through either brute-force methods or by exploiting unchanged default credentials. Attack motivations range from vandalism, the outright disabling of the infected devices, to various forms of financial gain by, for example, selling DDoS services to third parties or dropping cryptocurrency miners onto the infected devices.

Persistence

One of the new threats in the context of (IoT) malware has been the emergence of persistent malware. Persistent malware, contrary to its non-persistent counterpart, typically remains on the infected device even after remediation steps such as rebooting the device have been taken or otherwise frustrates removal of the infection. As such, the efforts required to properly clean a device of persistent malware are significantly much more difficult without the intervention of experts. VPNFilter and QSnatch represent two of the more widely spread persistent malwares, although persistent variations of other well-known malware families have also been identified. The increased difficulty of cleaning persistent malware from an infected device, coupled with its potential to render devices unusable even after factory resets, further emphasize the need for mechanisms to detect and prevent the spread of malware.

2.2. Malware Mitigation

The necessity to coordinate attacks provides an opportunity for the detection of botnets through the analysis of the traffic sent across these channels or through the identification of C&C hosts (Zeidanloo et al., 2010). However, these efforts can be hampered by resorting to alternative communication channels such as social media platforms in order to avoid detection across traditional channels, or by resorting to encryption techniques.

Feily et al. (2009) conduct an early survey of botnets and botnet detection techniques. The authors distinguish four types of detection techniques; signature-based detection, anomaly-based detection, DNS-based detection, and Mining-based detection. Signature-based detection techniques monitor network traffic for signatures and behaviour of known botnets. This, however, means the method is not suitable for the detection of unknown botnets. Anomaly-based detection techniques attempt to detect botnets based on network traffic anomalies such as high-latency traffic, high-volume traffic, or traffic on unusual ports. DNS-based detection techniques are similar to anomaly-based detection techniques in that they monitor for unusual DNS traffic. Mining-based techniques are primarily used to detect communications between the C&C server(s) and the botnet. Since C&C communications typically employ widely used protocols and display few traffic abnormalities, anomaly-based detection is generally not possible. Clustering and machine-learning techniques, among others, may then be used to attempt to identify distinguishing features of these communications in large datasets.

Feily et al. identify DNS-based techniques as highly promising; a DNS BlockList (DNSBL) approach proposed by Ramachandran et al. (2006) is the only surveyed method that allows for real-time detection of botnet activity, alongside its ability to detect encrypted bot communications. Both the DNS-based and mining-based techniques are further noted as having low false-positive rates and being agnostic to the botnet protocol (whereas most of the explored techniques function only on specific C&C architectures). The following paragraphs provide a short overview of the domain name system and explore the use of DNS-based methods for botnet detection and mitigation.

2.2.1. The Domain Name System

Under normal circumstances, the domain name system is used to locate the IP addresses of network resources a client wants to access; providing the translation mechanism between a domain (e.g. www.google.com) and the IP address(es) at which its service or resources are deployed. Figure 2.3 provides a simplified overview of a DNS query being processed by the domain name system.

A client machine (such as a personal computer) wants to navigate to a certain website, but it does not know the IP address of the website's web server. The client queries a DNS server for the IP address associated with the website. These DNS servers typically reside with the internet service provider, although clients can configure private DNS servers or use third party services for DNS resolution. DNS queries originating from a client initially pass through a recursive resolver (*recursor*) which attempts to locate the domain in its local, short-term storage (cache). If there are no cache hits for the requested domain, the recursor queries the DNS root server. The root server refers the recursor to a Top-Level Domain (TLD) nameserver based on the extension of the domain (e.g. '.com'). The TLD nameserver then redirects the recursor to an authoritative nameserver, from which it will typically receive the IP address of the domain requested by the client. The recursor returns the IP address to the client machine which can now request the resources from the web server.

A number of different types of records may be returned by a DNS recursor or other part of the DNS infrastructure to a requestor. Table 2.1 lists the most commonly encountered types of DNS records and their purpose in the infrastructure.

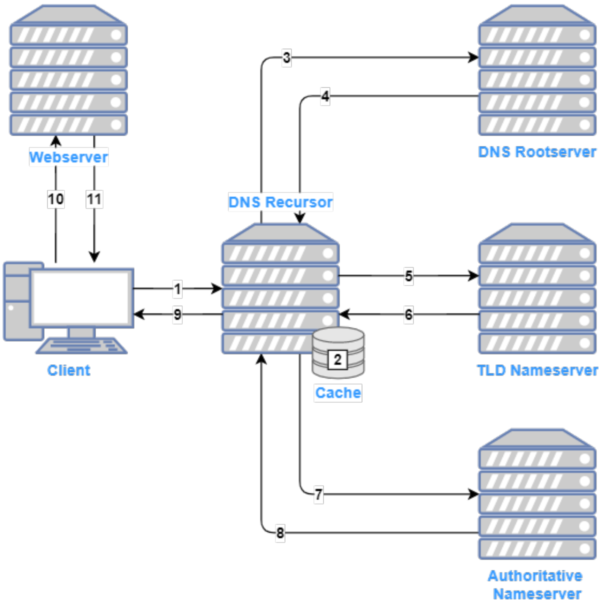


Figure 2.3: Simplified visualization of domain lookup using the domain name system.

As mentioned in section 2.1, malware abuses the domain name system primarily in order to identify and contact command and control servers. The use of DNS has obvious advantages over using IP addresses hardcoded within the malware binaries, as it prevents a loss of control in case of errors or elimination of the hardcoded address(es). DNS can also help mask the identity of the attacker, for example by employing proxy servers or by consistently changing the IP addresses of command and control servers. One of the major reasons why many malware families abuse the domain name system instead of other systems or protocols such as HTTPS or IRC is the limited control exercised over DNS traffic by network operators. Even under restricting circumstances, the relatively unrestricted use of DNS may be required for systems to function properly.

2.2.2. DNS-based Mitigation Techniques

Alieyan et al. (2017) survey the use of DNS-based botnet detection techniques and present a taxonomy that distinguishes a number of techniques. At the highest level, the researchers distinguish honeynet-based systems and intrusion detection systems (IDS). Honeynet approaches emulate known software and network vulnerabilities in order to provoke infection by botnets. These self-contained networks require minimal resources to set up, but have significant drawbacks in their limited scalability and interaction with malicious activities. As such, honeynets are primarily employed to recognize features and mechanisms of botnets but seldom to actively detect and mitigate malicious activity. Figure 2.4 presents the taxonomy of DNS-based intrusion detection systems constructed by Alieyan et al., which will be explored in further detail.

Table 2.1: Commonly found DNS record types

Type	Name	Description
A	IPv4 address	One or more IPv4 addresses (e.g. 172.16.254.1) associated with the requested host.
AAAA	IPv6 address	One or more IPv6 addresses (e.g. 2001:0db8:85a3:0000:0000:8a2e:0370:7334) associated with the requested host.
CNAME	Canonical Name	Aliasing of a domain to another domain, for example a redirection of requests to 'www.website.com' to 'website.com'.
MX	Mail Exchange	Used to identify the mail server to which an email directed at a certain domain must be delivered.
PTR	Pointer	Primarily used for reversed mapping, identifying a domain name by an IP address, such as: '45.34.23.12.in-addr.arpa'.
SRV	Service Location	Used to identify the locations of services such as the SIP-servers commonly used in VoIP applications.
TXT	Text Record	Textual information about a domain used for a variety of purposes such as ownership validation.

Intrusion Detection Systems

IDS are can be classified into one of either signature-/behaviour-based techniques, or anomaly-based techniques. Signature-based IDS detect known bots through signature matching, or by employing blacklists such as the technique proposed by Ramachandran et al. (2006) which contain the IP addresses of server machines associated with malicious activities. Signatures can also be based on code snippets or other defining characteristics of the malicious software. Antonakakis, Perdisci, et al. (2010) propose a dynamic reputation system that can be used to gauge the legitimacy of domains targeted by DNS queries. A major limitation associated with signature-based approaches is the need to maintain an up-to-date database of malicious addresses or other signature types, and the relative ease with which such detection techniques can be evaded (for example; by creating new versions of a malware with new signatures). Anomaly-based approaches are defined along the same lines presented by Feily et al. (2009); systems which attempt to detect botnets based on unusual traffic characteristics.

Anomaly-based Techniques

Anomaly-based detection techniques are further subdivided into host-based techniques and network-based techniques. Host-based techniques monitor and analyze system processes locally on each individual host. Examples of host-based techniques are firewalls, malware scanners, and more intricate technologies or other means of securing the host. Host-based techniques that operate on DNS are rare due to the availability of more comprehensive and powerful tools that can operate at level of individual hosts. However, host-based techniques experience severe limitations in scalability and are restricted to those systems that are being actively monitored. Thus, in order to achieve a level of mitigation across a wider network would require the installation of monitoring tools on a large number of hosts and mechanisms to arrange collaboration between these installations. Network-based techniques monitor activity at the network level in order to identify botnets, thus alleviating the scalability issues associated with host-based techniques.

Network-based Techniques

Network-based detection techniques are characterised as either active or passive techniques. Active DNS techniques utilize specially crafted packets that are injected into the network of interest. By capturing and analyzing the responses to these packets evidence may be gathered that hints at the presence of a botnet, for example by leveraging DNS probing methods to pinpoint malicious domains. Otgonbold (2014) employ active probing techniques to detect malicious fast-flux domains (domains whose IP association is cycled at a high rate among a network of compromised hosts). Ma et al. (2015) employ active probing of DNS caches to estimate malicious activity within networks. Nevertheless, actively monitoring and tracking malicious domain names (which may be removed or expire over time) in large, distributed networks requires a large amount of resources. Passive DNS techniques - rather than attempting to track malicious domains - sniff out traces of botnet activity among DNS queries.

Passive DNS Techniques

Passive DNS techniques aim to identify traffic originating from botnets based on anomalies within DNS queries. The DNS protocol is one of the most widely used protocols for C&C communication and the botnets that operate

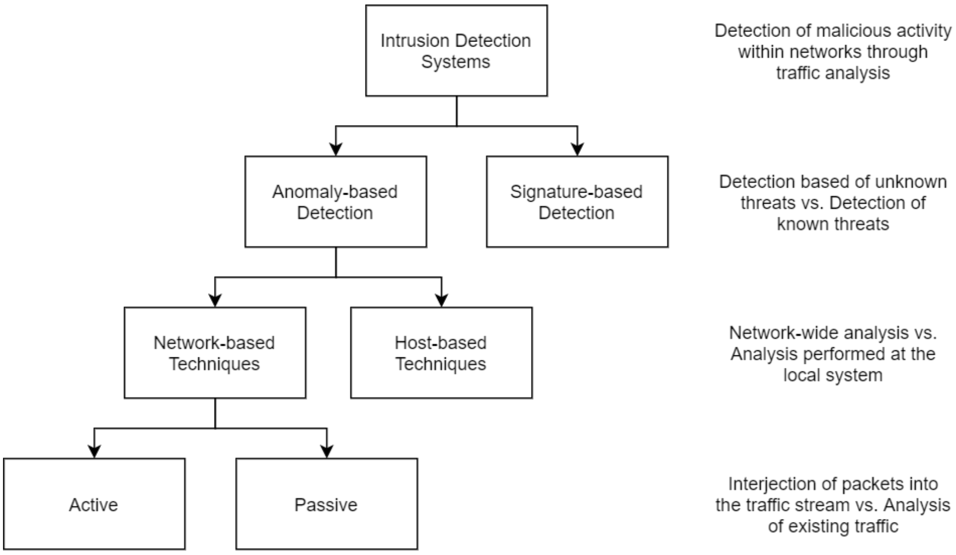

Intrusion Detection Systems
Detection of malicious activity within networks through traffic analysis
Anomaly-based Detection
Signature-based Detection
Detection based of unknown threats vs. Detection of known threats
Network-based Techniques
Host-based Techniques
Network-wide analysis vs. Analysis performed at the local system
Active
Passive
Interjection of packets into the traffic stream vs. Analysis of existing traffic

Figure 2.4: Taxonomy of DNS-based intrusion detection methods.

through the domain name system use easily-processed domain names to locate the C&C servers, or to transfer stolen data from compromised hosts. A variety of methods exist to analyze DNS traffic, of which the researchers distinguish six: graph theory methods, statistical methods, entropy-based methods, clustering methods, neural networks, and decision trees.

K. Wang et al. (2011) propose a fuzzy pattern recognition algorithm for the detection of botnets. The algorithm operates on the basis of the distribution of DNS query packets in its first phase (to detect inactive bots attempting to connect to command and control servers) and network traffic characteristics (to detect active bots based on ongoing malicious activity). The need to frequently connect with C&C servers, and the often very similar payload sizes of these communication packets, allow for reasonably high accuracy and low false-positive rates. Nevertheless, the reliance on such obvious interaction patterns put to question the effectiveness of these relatively simple techniques in detecting more advanced malware types.

Bilge et al. (2011) train a classifier on fifteen features extracted from RNDS records; a combination of time- and answer-based features, features derived from the records time to live, and the characters and length of domain names. Despite a relatively high false positive rate, the authors prove the solution can be employed feasibly and scalably at the site of a (small) internet service provider as a real-time early warning system.

J. Lee & Lee (2014) employ a clustering and graph-based method dubbed ‘GMAD’, graph-based malware activity detector, to detect infected clients and malicious domain names through sequences of DNS queries. Promisingly, GMAD is capable of detecting malicious activity - in spite of the use of evasion techniques - and with exceptionally low false-positive rates.

Shi et al. (2018) employ a neural network for the detection of malicious domains, using features similar to those used in the EXPOSURE method presented by Bilge et al.. Shi et al. achieve detection and accuracy rates of approximately 96% under optimal circumstances, outperforming classifiers based on more traditional techniques such as logistic regression and support vector machines.

Response Policy Zones

The response policy zone (RPZ) is a method used to deny access to malicious or bad domains. RPZs employ customized policies so that DNS recursors may return modified responses to DNS queries if the request is deemed inappropriate (Vixie & Schryver, 2017). RPZ responses may therefore be employed to return modified CNAME records (a fake alias) in order to redirect traffic (commonly applied in walled garden environments, where the administrator wants to put restrictions on accessing web content), fake NXDOMAIN records (indicating the domain does not exist), or explicitly exempt specific records from policy effects (Connery, 2013). The RPZ standard provides a number of benefits as well as a number of foreseeable drawbacks, with censorship imposed by governments or other powerful actors as one of the most commonly cited problems of employing response policy zones.

2.2.3. DNS Traffic Security

The fact that DNS queries are sent across the internet in an unencrypted manner introduces the risk that these packets get snooped on or manipulated by a third party; resulting in privacy and security deficiencies. Efforts to resolve these deficiencies by encrypting DNS traffic nevertheless frustrate the analysis of DNS traffic and, thus, efforts to identify malicious queries. Over the last years two notable techniques have been introduced and increasingly adopted to standardise the encryption of DNS traffic; DNSSEC and DNS over HTTPS/TLS.

DNSSEC

While providing opportunities for malicious activity detection, the unencrypted nature of the UDP protocol used to deliver DNS queries makes the system inherently vulnerable to abuse beyond the mere facilitation of communication between malicious hosts. Malicious actors may compromise the unencrypted UDP packets through the insertion of malicious content or by otherwise modifying the payload they carry (Hinchliffe, 2019), through man-in-the-middle attacks (Libfeld, n.d.), or by inserting illegitimate records into legitimate DNS servers (*What is a DNS Hijacking: Redirection Attacks Explained*, 2019). In order to tackle this problem, the Domain Name Security Extension (DNSSEC) has been introduced to enhance the security of DNS traffic. DNSSEC adds authentication and integrity mechanisms relying on the addition of digital signatures through public and private cryptographic keys, thereby allowing clients to verify the integrity of a response (although it does not provide encryption mechanisms).

However, the inclusion of the digital signatures, alongside the introduction of additional record types, significantly increases the overhead of extended DNS records. DNSSEC packets may be up to four times as large as the packets traditionally used to transmit DNS data (512 bytes for classical IPv4 packets, recently extended to 1280 bytes to account for IPv6 support, while 2000 bytes or more are common for DNSSEC packets), leading to the potential fragmentation of DNSSEC records as they are sent across networks and significantly increasing the complexity of DNS services (Kim & Reeves, 2020). Due to this complexity of DNSSEC implementations relative to traditional domain name services and incompatibility issues that may lead to unexpected behaviour and instability (Institute, 2021).

A 2017 investigation into the adoption of DNSSEC among the .nl top-level domain indicates that, while DNSSEC adoption is growing at a significant pace, notable disparities exist between various industrial sectors (for example, only 6% of banking related domains had been signed in accordance with the standard, in contrast with 64% of domains related to internet infrastructure). Statistics published by the SIDN (Stichting Internet Domeinregistratie Nederland) in indicate that as of June 2021 approximately 73% of DNS queries originated from DNSSEC-enabled resolvers, while roughly 56% of domains employ DNS signatures (*.nl statistieken: SIDN Labs*, 2021).

DNS over HTTPS/TLS

The DNS over HTTPS (DoH) and DNS over TLS (DoT) protocols are similar to DNSSEC in their intent, but can be considered stronger or more strict measures as they involve the actual encryption of DNS traffic. DoH and DoT are - much like DNSSEC - intended to alleviate privacy issues and improve overall security by preventing access to, and manipulation of, DNS traffic through man-in-the-middle attacks. They achieves this by encrypting DNS requests between the client and DNS resolver, thereby frustrating attempts to intercept and eavesdrop on DNS traffic. DoH support has seen relatively widespread adoption by various public DNS servers, among open-source DNS resolver solutions, and in major web browsers. While the intended implementation is that of end-to-end encryption, DoH is often implemented as hop-to-hop encryption instead.

While the use of DNS over HTTPS does not directly impact the functioning of DNS-based malware mitigation solutions hosted at the site of the DNS resolver (which may examine the decrypted traffic), it presents an issue for solutions that monitor traffic at sites other than the intended resolver. The Dutch National Cyber Security Center concludes that the DNS-over-HTTPS and DNS-over-TLS protocols may severely impact the functionality of centralised DNS monitoring (NCSC, 2019), as parties such as Mozilla and Google are rapidly rolling out support across their software. The Godlua malware identified in 2019 is one of the first widely known malwares to employ DoH to hide its traffic with command-and-control servers, hinting at the awareness of threat actors of the potential for unscrutinized botnet control and data exfiltration using these protocols.

2.3. Adoption of Security Measures

Despite the potential improvements to home security and the relatively small effort required for end-users to adopt centralised security measures their adoption remains low (Antonakakis, Dagon, et al., 2010; KPN, 2020). Key to understanding why this is the case, is to identify the factors that encourage or inhibit the adoption of security measures. Various researchers investigate the determinants for compliance in security behaviour, predominantly in enterprise or organizational environments (Bulgurcu et al., 2010; Safa et al., 2016; Y. Li & Siponen, 2011). Comparatively little work has been done to identify determinants of home-user security behaviour, especially in the context of centralised security measures.

The following sections dive into a number of theories that have been widely used to study security behaviour and the adoption of cyber security measures. Sections 2.3.1 and 2.3.2 explore two prominent frameworks for the acceptance and use of technology; the Technology Acceptance Model (TAM) and its successor the Unified Theory for the Acceptance and Use of Technology (UTAUT). Sections 2.3.3 and 2.3.4 go into the Health Belief Model (HBM) and Protection Motivation Theory (PMT) respectively. Section 2.3.5 touches upon the Technology Threat Avoidance Theory (TTAT).

2.3.1. Technology Acceptance Model

The technology acceptance model (TAM, Figure 2.5) is perhaps the longest standing and most straightforward model that attempts to explain the adoption of technology; theorizing that a technology's perceived usefulness, coupled with its perceived ease of use, are key determinants for the behavioural intention to use, and the actual use of novel technologies (Davis, 1985). TAM finds its origins in the field of psychology, in the Theory of Reasoned Action (TRA) and Theory of Planned Behaviour (TPB); the latter being an extension of the former. The theory of reasoned action built upon the principle that individuals tend to make rational, systematic use of information to inform their behaviour; the theory of planned behaviour identified the importance of perceived behavioural control to explain discrepancies between intention and actual behaviour (Ajzen, 1991). Despite its wide acceptance and use, TPB has notable limitations due to its exclusion of subconscious motives and demographic variables as predictors for behaviour. Consolidation of the variables included in the theory of planned behaviour - largely into two new variables: perceived usefulness and perceived ease of use - led to the formulation of TAM in 1985.

Y. Lee et al. (2003) conduct an extensive analysis of research conducted using the technology acceptance model, finding almost universal proof for the significance of perceived usefulness on the behaviour intention to use a system. The relationship between perceived ease of use and behavioural intention is notably less universal, although the authors note that several of the reviewed works provide an explanation for this lack of significance. For example, reality within organizations may simply not permit ease-of-use as a significant determinant (hierarchy might dictate the use of certain technologies regardless of user friendliness), or technologies may be inherently easy to use thus eliminating the parameter during decision-making. Marangunić & Granić (2015) note several avenues for further exploration of the TAM model; the inclusion of additional variables and an investigation into the relationship between actual use and objective outcomes. While the TAM is extensively used to study adoption within a variety of technological contexts, limited research efforts examine the application of TAM to security behaviour.

Kumar et al. (2008) extend the technology acceptance model with a large number of constructs, such as security awareness, computer anxiety, and concern for information privacy. The authors identify the necessity of greater adoption of security measures in the face of attacks on critical infrastructure, and find that computer anxiety and perceived usefulness are significant determinants for firewall use. Nevertheless, they also note a likely dependency of these significance of these constructs on the nature of the technological solution.

Jones et al. (2010) customize the technology acceptance model to study employee acceptance of security measures, adding the subjective norm as an antecedent for behavioural intention and management support as a moderating variable. In this context, neither perceived usefulness nor perceived ease of use were deemed to have a strong impact on security compliance. P. A. Wang (2010) adapts TAM to address security measure acceptance from a knowledge perspective, defining information, awareness, and experience as antecedents to attitude toward using and intention to use. Perhaps to little surprise, the researchers find that computer knowledge, experience, and awareness of security measures are positively associated with the intention to use security measures.

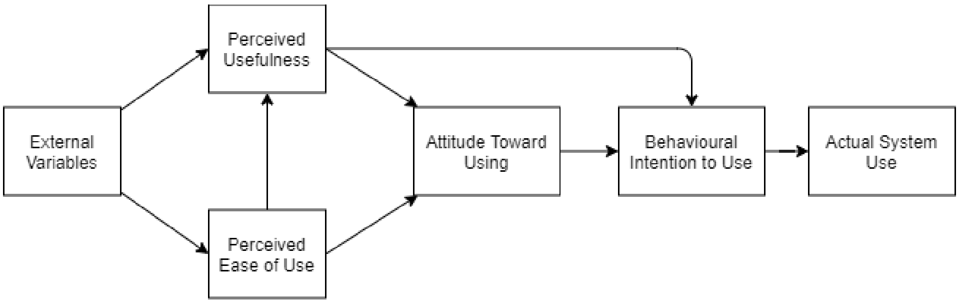


Figure 2.5: The Technology Acceptance Model.

2.3.2. Unified Theory of Acceptance and Use of Technology.

The unified theory of acceptance and use of technology (Fig 2.6) seeks to remedy some of the issues facing TAM, such as its exclusion of cognitive factors in the explanation of adoptive behaviour (Ng et al., 2009). By consolidating a number of prominent social, cognitive, and behavioural theories UTAUT provides a comprehensive overview of the factors that influence and moderate acceptance behaviour. The framework defines performance expectancy, effort expectancy, social influence, and facilitating conditions as antecedents for behavioural intention and actual usage behaviour. Gender, age, experience, and voluntariness of use as identified as important moderating variables for these relationships.

Like the technology acceptance model, UTAUT is widely applied in a number of technological contexts including those contexts where security and security perception are considered important reasons for adoption. Despite its wide application and strong predictive power, UTAUT has seen little application to explain security behaviour or the adoption of security measures (Venkatesh et al., 2016). While the framework has seen little application in the context of cyber security, the variables embedded in the framework are found as frequent additions to theories used in other studies.

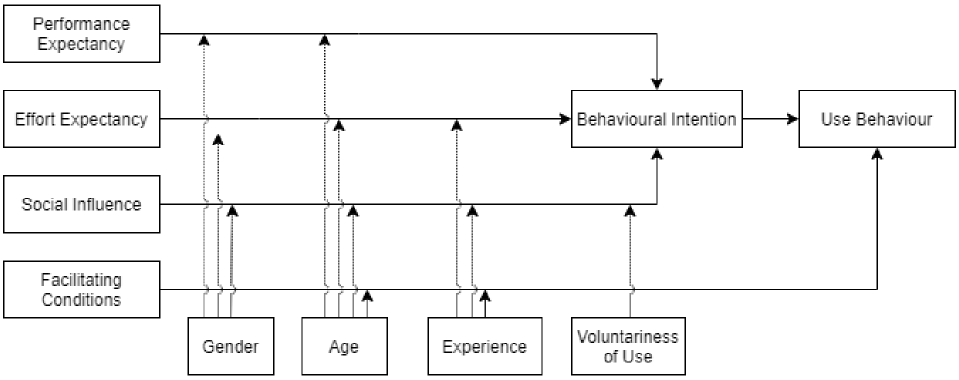


Figure 2.6: The Unified Theory of Acceptance and Use of Technology.

2.3.3. Health Belief Model

Several researchers note the parallels between the use of protective technology and preventive actions in healthcare. Figure 2.7 depicts the health belief model. A number of research efforts have examined the applicability of the health belief model in predicting security behaviour (Claar, 2011; Ng et al., 2009; Dodel & Mesch, 2017). While theories such as TAM and TPB have previously been used to study the adoption of computer security measures, they fail to capture the significant difference between positive technologies, which are used for their utility benefits, and protective technologies, which are instead used to prevent harm (Dinev & Hu, 2007).

The health belief model stipulates that perceptions of a threat, represented by an individual's perceived susceptibility to the threat and its perceived severity, together with an intrinsic motivation to mitigate the threat, the perceived benefits and barriers associated with protective actions, and external cues to action motivate an individual to take a protective action. These perceptions and the intrinsic motivation are, in turn, considered to be the result of a number of demographic and psychological variables such as personality traits and biological determinants.

Ng et al. (2009) apply the HBM to study email security behaviour, finding that perceived susceptibility, perceived benefits, and self-efficacy are determinants for security behaviour, consistent with the protective nature of security technologies and previous studies on intention and behaviour. Perceived barriers, cues to action, and perceived severity were not found to be significant reasons to deter from practicing computer security. Although perceived severity was not found to be a significant predictor of security behaviour, the authors note its moderating effects on the other variables. Nevertheless, the study's context is limited, both in terms of study population and the researched practices of email security.

Claar (2011) adapt the HBM in a cross-sectional study of home user personal computer security behaviour. The authors find that perceived vulnerability, perceived barriers, and self-efficacy are the primary determinants for security usage. Alongside these variables, the researchers identify the interaction between prior experience and perceived severity, and prior experience and self-efficacy as strong indicators of security behaviour. Nevertheless, a large number of the relationships hypothesized in the health belief model yield no significance.

Dodel & Mesch (2017) similarly adapt the HBM to study the antecedents of preventive behaviour in the context of cyber security, specifically the use of anti-virus software, firewalls, and file scanners. The authors find that while users are generally willing to put time and effort into security practices, few mentioned anti-malware as effective measures against cyber threats. In line with earlier findings, perception of severity and susceptibility, alongside the age and activity of internet users were found to be significant determinants. Moreover, self-efficacy (or the perception thereof) in protecting devices and the willingness to go through the effort of actually protecting them were identified as critical in the adoption of antivirus measures.

2.3.4. Protection Motivation Theory

The protection motivation theory was first formulated by WRogers in 1975, but it's most common form is a revised version developed in 1983. PMT has been widely used study the adoption of protective measures in psychology and healthcare, as well as in information security (Woon et al., 2005; Johnston & Warkentin, 2010; Vance et al., 2012; Crossler & Bélanger, 2014; Hanus & Wu, 2016; Martens et al., 2019). Similar to the health belief model, PMT

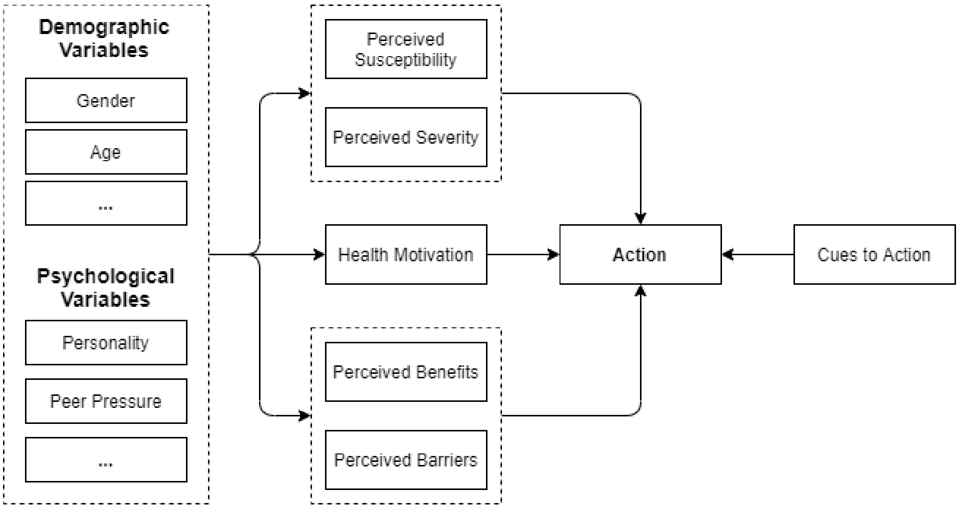


Figure 2.7: The Health Belief Model.

specifically accounts for the protective nature of security applications (vis-a-vis there mere adoption of technologies for utility), and the fear appeals that play a significant role in motivation towards protective behaviour. Figure 2.8 presents a model of the cognitive mediation process in the revised protection motivation theory.

Central to PMT are the concepts of *threat appraisal* (an individual's perception of a threat) and *coping appraisal* (the ability to cope with the threat). Threat appraisal is determined by its severity and the individual's vulnerability towards it. Coping appraisal is determined by the response efficacy (the perceived adequacy and benefits of implementing the behaviour), self-efficacy (an individual's ability to implement the behaviour), and response costs (drawbacks associated with implementing the behaviour). Depending on the cognitive mediation between these appraisals, an individual takes a maladaptive coping response (refraining from implementing the protective behaviour) or an adaptive coping response (implementing the protective behaviour).

Woon et al. (2005) employ PMT in a survey study of 189 home users' adoption of wireless network security measures. The authors find evidence to support perceived severity, response efficacy, self-efficacy, and response cost as significant predictors. Users' perceived vulnerability to a threat was not found to be a significant predictor, suggesting that while individuals may consider themselves as unlikely victims of security breaches, the potential damage caused by a breach may persuade them to adopt protective behaviour. The authors note that self-efficacy and response-efficacy are the predominant concerns among users that appealed to others to help them enable the measures, indicating the importance of social factors in their adoption.

Johnston & Warkentin (2010) examine the adoption of anti-spyware software among college students, identifying response efficacy, self-efficacy, and social influence as strong predictors for the intention to use such software.

Vance et al. (2012) examine compliance with information security behaviour using PMT and habit theory: despite several efforts have been made to study the application of PMT to IS security, they have rarely attempted to capture the downsides of security behaviour and the effect of prior experience. Habit towards compliance with information security policies was found to have a significant impact on all variables in PMT; although it should be noted that this may partially or fully explained by the enterprise-context of the research. Additionally, threat severity and vulnerability were found to have significant impacts on security behaviour; stressing the importance of creating awareness among users.

Crossler & Bélanger (2014) conduct a survey on security practices among 300 citizens, identifying severity, response efficacy, and self-efficacy as significant variables. The authors' reliance on self-reporting behaviour is nevertheless identified as a limitation of the study; noting that social desirability bias may result in lower support for the significant hypotheses in reality. Thus, the ability to measure not only intention through self-reporting or other methods, but also to measure actual security behaviour is noted as an important point of improvement.

Hanus & Wu (2016) investigate the impact of user's security awareness on desktop security behaviour using the protection motivation theory. The authors note the lack of research into home-user security behaviour, despite an increased vulnerability to threats. In line with previous findings, self-efficacy and response efficacy are identified as significant predictors for security behaviour, while response cost did not have a significant impact on security adoption (potentially due to the familiarity of respondents with the available security measures, and the fact that such protective software is often free or bundled with other products). While threat appraisal overall is found to not have a significant effect on behaviour, user awareness of both threats and countermeasures plays an important role in motivating the use of protective behaviour.

Martens et al. (2019) compare predictors for the intention to take security measures between malware threats and scams, finding significant differences between the predictive models of both contexts. Perceived severity and response efficacy were found to be highly significant predictors, in turn strongly influenced by participants' awareness of threats and countermeasures. Perceived vulnerability was found to be significant only in the case of malware, and self-efficacy was found to be significant only among those with a high degree of technical know-how.

A 2021 article by Haag et al. presents a comprehensive, systematic review of the application of PMT in the field of information security; examining 61 articles from 2005 to 2017 that empirically evaluate the model. The authors propose a number of recommendations for future applications of PMT in information security research in five areas. They propose that researcher measure the degree of concern experienced by subjects with regards to threats, as well as confidence in the relationship between coping behaviour and effective threat reduction. Six directions to personalize IS security threat messages are suggested in order to examine in greater detail the activation of beliefs in subjects. The researchers provide recommendations to study maladaptive coping with emotions such as fatalism and denialism - maladaptive coping modes were found to be included in only 3 out of 61 PMT-based studies - as a means to study the growing knowledge gap present in cyber security. Lastly, they suggest researchers study the impact of personality variables on coping behaviour.

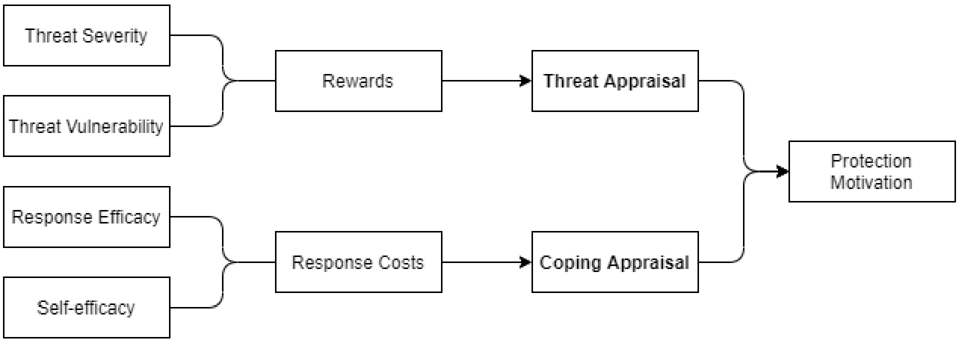


Figure 2.8: The cognitive mediation process of Protection Motivation Theory.

2.3.5. Technology Threat Avoidance Theory

The technology threat avoidance theory (Figure 2.9) presents an integrated view of the health belief model, the protection motivation theory, and risk analysis research (Liang & Xue, 2009). Like the theories the model is rooted in, it asserts that motivation to engage in protective behaviour (denoted as avoidance motivation) results from the interplay between threat perception, the efficacy and costs of the proposed safeguarding measure, and an individual's efficacy in enacting the behaviour. In their initial evaluation of the model in the use of anti-spyware measures among a population of college students, Liang & Xue (2010), find evidence for the validity of all core constructs in their explanation of avoidance motivation and avoidance behaviour. The model accounts for 56% of the variance in avoidance motivation and 21% of the variance in avoidance behaviour.

Arachchilage & Love (2014) examine the applicability of the TTAT on computer users' thwarting of phishing attacks, limiting the model to definitions of procedural and conceptual knowledge, their effect on self-efficacy, and its effect on avoidance behaviour. The authors employ a questionnaire of items combined from previous research efforts using a 5-point Likert scale to capture responses from 161 undergraduates at two universities, with a predominantly high exposure to internet activity. Like previous studies, Arachchilage & Love find that self-efficacy is a strong predictor for avoidance motivation (and by extension; avoidance behaviour).

Young et al. (2016) apply the technology threat avoidance theory on a sample of 486 computer users, revealing safeguard effectiveness, safeguard cost, and self-efficacy as robust motivators for avoidance behaviour across contexts. Threat perception was found to be overall a less stable predictor of such behaviour under changing circumstances. While their results underwrite the validity of TTAT as a framework for user behaviour surrounding cyber security behaviour and malware in particular, the authors note the necessity of examining further predictors such as trust and risk propensity, and the effects of social influence.

The extensive body of research into the applicability of protection motivation theory in predicting both information security behaviour in general, as well as the adoption - or intention to adopt - malware countermeasures, provides a solid starting point for the identification of determinants for centralised malware mitigation measures. In line with the reviewed literature, a model is proposed to examine the significance of well-tested constructs in predicting the adoption of a centralised, DNS-based malware mitigation service.

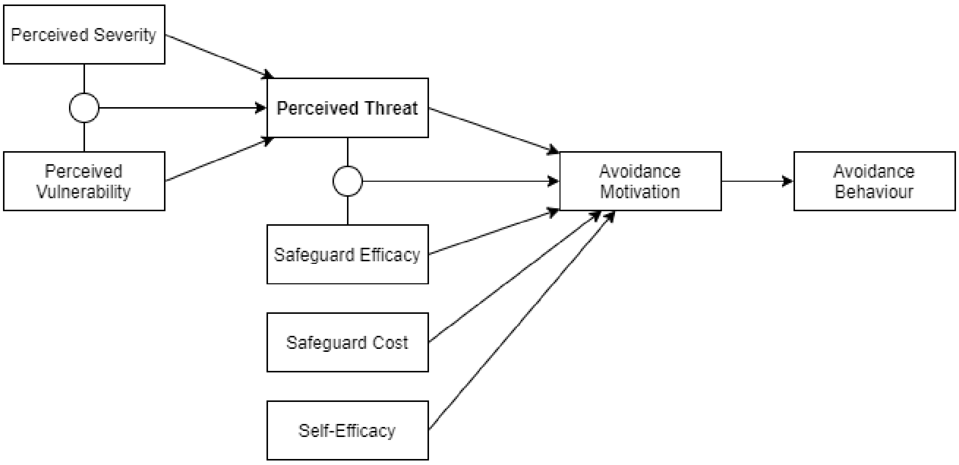


Figure 2.9: The Technology Threat Avoidance Theory as first used by Liang & Xue (2010).

2.4. Synthesized Model

Based on the reviewed literature, a number of consistently relevant constructs in the adoption of security measures can be identified. While the direct effects of threat severity and threat vulnerability in explaining coping behaviour are situational, especially among non-IT professionals (Kumar et al., 2008), several studies stress their relevance (Martens et al., 2019; Young et al., 2016; Dodel & Mesch, 2017) alongside the mere awareness of both threats and security measures (P. A. Wang, 2010; Vance et al., 2012; Hanus & Wu, 2016). Response efficacy and self-efficacy are almost universally identified as key determinants for security behaviour (Woon et al., 2005; Kumar et al., 2008; Claar, 2011; Crossler & Bélanger, 2014; Arachchilage & Love, 2014; Dodel & Mesch, 2017). The perceived costs of security measures are typically not a relevant factor in environments where compliance is expected or required, or when costs are low (Ng et al., 2009; Hanus & Wu, 2016; Dodel & Mesch, 2017). However, studies specifically aimed at home-user security behaviour indicate its relevance within that context (Woon et al., 2005; Kumar et al., 2008). Moreover, a number of studies identify the moderating effects of demographic variables such as gender, age, and experience (P. A. Wang, 2010; Claar, 2011; Dodel & Mesch, 2017) and the relevance of social influence in the adoption and implementation of security measures (Woon et al., 2005; Kumar et al., 2008; Johnston & Warkentin, 2010; Young et al., 2016).

The major shared or overlapping constructs of various theoretical frameworks such as threat perception (severity and vulnerability) and response perception (efficacy and cost) are expected to provide situational but potentially significant predictive power. Self-efficacy is a significant predictor among almost all works, and several researchers denote the significance of - or need to study - the effect of external influence (typically referred to as social influence or subjective norm). Table 2.2 provides a definition of the constructs included in the conceptual model.

Table 2.2: Definitions of the constructs included in the conceptual model.

Construct	**Definition**
Perceived Severity (***PS***)	An individual's assessment of the severity of the consequences of a threat event (Ifinedo, 2012).
Perceived Vulnerability (***PV***)	An individual's assessment of the probability of being exposed to a threat (Maddux & Rogers, 1983).
Perceived Response Efficacy (***PRE***)	Beliefs about whether the response will be effective in reducing the threat (Herath & Rao, 2009).
Perceived Response Costs (***PRC***)	Beliefs about costs (e.g., money, time, effort, side-effects) associated with taking the suggested coping response (Y. Lee & Larsen, 2009).
Perceived Self-Efficacy (***PSE***)	The belief that one is or is not capable of performing a coping behavior (Y. Lee & Larsen, 2009).
Subjective Norm (***SN***)	The perceived need to perform or not perform an action derived from an individual's perception of the views held by others (Tsai et al., 2016).

2.4.1. Intention to Adopt Centralised Security

Based on the conceptual model a number of hypotheses can be formulated to predict and test the relationship between the established PMT constructs and the willingness to adopt centralised security measures. The first two hypothesized relationships concern how the end-users' threat perception is expected to affect their adoption of mitigation measures.

Participants who perceive the threat of malware infections to be greater are more likely to adopt the mitigation measure than those that do not. This hypothesis stems from the assumption that between two individuals, the one who perceives that a threat has a greater potential for damage will be more likely to attempt to mitigate the threat event as the perceived losses associated with a successful threat event are greater. As participants view themselves or their devices as increasingly vulnerable to threats they are expected to engage in protective behaviour to reduce their exposure to these threats. As such, those participants with a greater perception of vulnerability are expected to adopt the mitigation measure more frequently.

Hypothesis 1 *Perceived severity positively affects the intention to adopt the malwarefilter.*

Hypothesis 2 *Perceived vulnerability positively affects the intention to adopt the malwarefilter.*

The third, fourth, and fifth hypotheses concern how end-users' perception of the available response affects its adoption. Participants will be more likely to adopt the response as they perceive the measure to be more effective at mitigating potential threats, thus making its adoption more attractive. Conversely, if an individual perceives that employing the response comes at a great cost, they will be less likely to adopt it. After all, the greater cost of the response would offset the potential benefits it may bring to its user.

Individuals with a greater perception of self-efficacy are expected to be more likely to adopt the response as a result of these participants encountering fewer barriers both in terms of effort expectation and the actual implementation. Conversely, participants with a low perception of self-efficacy will have more difficulty overcoming the thresholds associated with adopting the protective measure.

Hypothesis 3 *Perceived response efficacy positively affects the intention to adopt the malwarefilter.*

Hypothesis 4 *Perceived response costs negatively affects the intention to adopt the malwarefilter.*

Hypothesis 5 *Perceived self-efficacy positively affects the intention to adopt the malwarefilter.*

Finally, the sixth hypothesis concerns the effect of the end-users' beliefs about their environment in the adoption of the malwarefilter. It builds on the expectation that participants who experience a higher degree of external influence are more likely to adopt the response measure. This influence may exist in the form of heightened awareness of the topic due to its importance among their social circle, in their professional environment, or the perceived need to comply with social norms concerning the need for - or use of - measures to protect one's devices.

Hypothesis 6 *Subjective norm positively affects the intention to adopt the malwarefilter.*

2.4.2. Maladaptive Coping with Emotions

In line with the recommendations of Haag et al. (2021), and given the notable absence of maladaptive coping modes in earlier works, the conceptual model includes not only the commonly used predictors or core constructs of the PMT model, but also the dependent variable of maladaptive coping with emotions. This construct is based on the notions of maladaptive coping with emotion as suggested by Haag et al. (2021); denialism, fatalism, and wishful thinking.

Threat appraisal, embodied by the constructs of perceived severity and perceived vulnerability, is hypothesized to positively influence maladaptive coping. That is; increasing perceptions of the potential damage a threat may do, and increasing perceptions of an individual's vulnerability to such threats, are expected to in turn increase the degree to which the individual engages in maladaptive coping with emotions. For example, individuals that consider themselves highly vulnerable to online threats may have the disposition that these threats are so grave that nothing can be done to mitigate them (i.e. engage in fatalistic emotions).

Hypothesis 7 *Perceived severity positively affects maladaptive coping with emotions.*

Hypothesis 8 *Perceived vulnerability positively affects maladaptive coping with emotions.*

Response efficacy is expected to negatively affect the degree to which individuals engage in maladaptive coping with emotions. Individuals that are presented with countermeasures that they consider to be effective at mitigating the threats they are faced with, it becomes less necessary to engage in maladaptive coping behaviours due to the availability of an effective coping response. Conversely, if such a response is considered to be costly (whether financially or due to other harmful side-effects), individuals might instead resort to maladaptive coping with emotions such as denialism ("I don't need to pay the costs for these measures since they will not be necessary anyway") or wishful thinking ("if only I did not have to think about these things").

Hypothesis 9 *Perceived response efficacy negatively affects maladaptive coping with emotions.*

Hypothesis 10 *Perceived response costs positively affects maladaptive coping with emotions.*

Perceived self-efficacy and subjected norm are both hypothesized to negatively affect engagement with maladaptive coping with emotions. End-users that consider themselves more capable of implementing security measures are expected to be less likely to engage in maladaptive coping with emotions as there is less need for denialism, fatalism, or wishful thinking if one considers themselves capable of implementing measures to prevent or mitigate threats. Similarly, individuals who perceive that topics such as online security are important to their environment might have the expectation that more is being done to prevent threats already and therefore may not need to resort to maladaptive coping with emotions.

Hypothesis 11 *Perceived self-efficacy negatively affects maladaptive coping with emotions.*

Hypothesis 12 *Subjective norm positively negatively affects maladaptive coping with emotions.*

2.4.3. Conceptual Model

Figure 2.10 presents the conceptual model that theorizes the predictors for the adoption of centralised malware mitigation services. The model is similar to the protection motivation and technology threat avoidance theories; restricting itself to those constructs that are either consistently significant or strong but situational predictors for protective behaviour. The significance of these constructs in predicting the adoption of centralised DNS-based malware mitigation measures and maladaptive coping with emotions is examined, which are represented by the dependent variables of 'Intention to Adopt' and 'Maladaptive Coping' respectively, and the expected effect direction of these relationships based on the hypotheses defined above are visualized through arrows and positive and negative sign symbols.

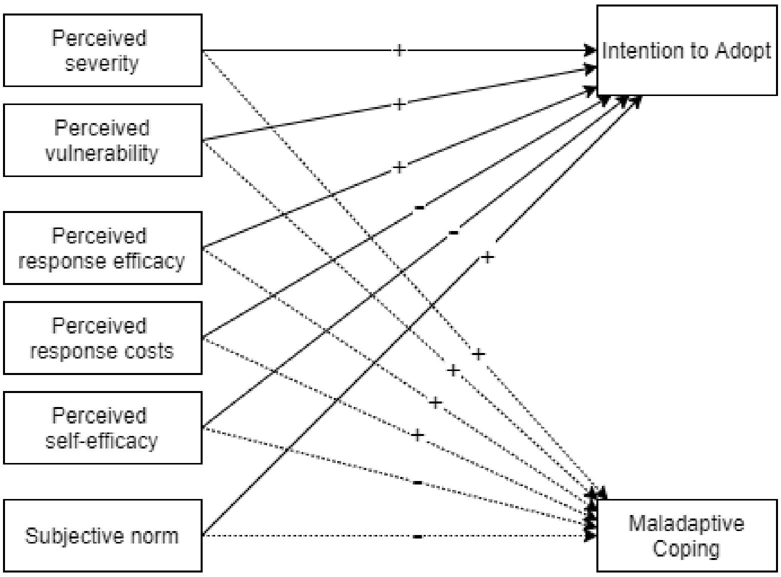


Figure 2.10: Conceptual framework synthesised from the examined literature.

3 Experiment Ecosystem

This section presents an analysis of the system and actors that defines the environment in which the study takes place. The most important actors and their role in the mitigation or propagation of cyber threats are explored. A schematic overview of the broader system and the system of interest are provided, the choice of study population is explored, and the manner in which the intervention affects the system is delineated.

3.1. The Actor Arena

The complexity of the problem, the mitigation of malware activity, is largely the result of the interactions between the various actors with conflicting interests and capabilities. As mentioned in chapter 1: end-users may be inclined to protect their devices but lack the (technical) skills required to do so, while device designers and manufacturers may lack the willingness to implement expensive security standards, and attackers adapt their methods and circumvent existing protections, and so forth. Table 3.1 lists the most prominent actors in the system, describes their roles within the system and their interests.

Table 3.1: The actors in the system and their interest

Actor	Description
End-user	The owner of the infected device, who may or may not be the primary target of malicious activity. Often, malware infections damage or otherwise impact the end-user even if they are not the primary target.
Internet service provider	The internet service providers are generally the parties responsible for providing network infrastructure to the general public. This infrastructure typically consists of both physical infrastructure and network services such as DNS services.
Security service provider	Security services come in a variety of forms. Security services are typically found either integrated into a device's operating system, or provided by a third party. Security services may provide local (host-based) protection through e.g. firewalls, or wider (network-based) protection such as DDoS mitigation.
Bot master	The bot master is the individual or group that exercises control over a botnet. Bot masters may target specific devices or device types, organisation, or networks for infection. Bot masters delegate the attack capabilities of the botnet either for financial gain (by selling attack services) or other reasons, typically through C&C servers that disseminate commands among the botnet.
Victim	In many cases, the intended victims of botnets are not the (owners of) the devices infected with the malware. DDoS attacks comprise the majority of cyber attacks, merely leveraging the infected devices to perform the attacks. However, malware may also be associated with ransomware, spyware, and other types of attacks where the infected device or its owner at the intended targets of an attack.
Customer	A large number of botnets are purpose-built not to inflict direct harm through the malware infection, but rather to leverage silently infected devices in primarily DDoS attacks. These attack services enabled by the botnet are typically sold through the deep web or even the surface web.
Government	The primary role of the government within the context is that of a regulatory or legislative body. As such, the government can impose (security) standards on devices brought to market, on network infrastructure, and define the manner in which service providers can or cannot attempt to mitigate malicious activity.
Device manufacturer	Device manufacturers can generally be assumed to have a vested interest in reducing the complexity and production cost of the devices they manufacture in order to appeal to a larger market or greater margins. High security standards may increase the complexity of design and testing processes which drive up manufacturing costs.

3.2. The System of Interest

Within the context of the wider system each actor provides interacts with one or more of the other actors in the system. Figure 3.1 defines the most important relationships between the various actors and their roles within the system. In this system, a smaller subsystem can be identified that is the subject of this study; the system of interest. While interactions outside the system of interest are not a subject of the study, they play a role in defining the boundaries within which one can approach a solution to the problem at hand, and may help explain behaviour within the system of interest.

Central to to the system of interest is the existence of malware botnets that can used to attack a variety of victims. These malware botnets are controlled by bot masters, who typically employ methods such as C&C servers to coordinate the actual devices comprising the botnet in order to spread and execute attacks. Within the system of interest lie the end-users and internet service providers as important actors. Internet service providers employ methods to mitigate the effects and spread of malware that operate across their networks. End-users are impacted by the spread of malware by being subjected to a malware infection, thereby involuntarily taking part in the botnet, but might also be the direct victims of attacks (as a result of an infection alone, as is the case with ransomware, or by being the target of DDoS attacks for example).

Outside the system of interest but within its wider context lie actors such as the government which is responsible for various regulations that directly impact the system as a whole, but which most prominently affect the operating requirements and principles of the internet service providers and define requirements and standards to which device manufacturers must or should adhere. Nevertheless, defining regulations and standards for novel internet connected devices remains an issue (Ahlmeyer & Chircu, 2016; Atlam & Wills, 2020).

Regulations aimed at the telecommunications market may impact the methods that can be employed to monitor and mitigate malicious activity across ISP's networks. For example, GDPR-like regulations can have severe consequences for the nature and amount of data that service providers are allowed to collect and the manner in which it may be analyzed. Restrictive regulations may impact the efficacy of methods employed to minimize abuse of network infrastructure and its connected devices, while overly liberal or lacking regulations may impact the privacy and interests of customers.

Regulations aimed at the device manufacturers and designers may impact the level of security and associated standards that devices must adhere to in order to be allowed onto the (consumer) market. By enforcing stricter or less strict regulations, government and legislative bodies may affect the ease with which devices can be exploited or should be resistant to exploitation by third parties, the configurability of devices, financial and non-financial expenses that must be made to test and ensure their adequacy and security, and more.

Two significant parties external to the system of interest that can alleviate the security issues associated with (IoT) devices are the device manufacturers, and the providers of security services. Device manufacturers' role within the system is defined by the design, production, and sale of potentially vulnerable devices to end-users. Although device manufacturers and designers are in a good position to maintain a high level of security both upon purchase of the device and throughout its lifetime by issuing software updates, they typically do not possess the willingness to implement measures that reduce battery life, increase time to market, or otherwise affect the complexity of a device (Yang et al., 2017).

Security service providers provide their services to a number of actors within the system depending on the strategic orientation of the organization. These services include measures aimed largely at the consumer market such as firewalls and malware scanners, and measures aimed at businesses and enterprise customers such as DDoS protection (CloudFlare is a prominent provider of such services) and malware reports (e.g. the ShadowServer reports that are used by the AbuseDesk).

Finally, bot masters, their victims, and the potential customers of botnet services lie just outside the system of interest, but are the primary external factors that influence or are influenced by their operation. While some botnets may be operated for the mere curiosity or malice of its controller, perhaps the most common reason to operate malware botnets are the financial gains that can be extracted either through the direct (threat of) damage to a victim, or the sale of the services enabled by the botnets (Z. Li et al., 2009).

The victims of botnet activity are disparate; from major private enterprises, to educational institutes, other public service providers, medium- and small businesses, and even individuals. While (potential) victims may find some measure of protection in various security measures provided by third parties or practices they employ themselves, perhaps greater potential for threat mitigation lies in controlling the spread of malware in the first place.

3.3. Population

The study aims to examine the adoption and efficacy of centralised, DNS-based mitigation services by end-users. The research takes places within the confines of the KPN ecosystem, which contains a variety of potential populations (customer groups) from which the subjects can be drawn; the consumer market, the business market, the wholesale market, and the mobile market. This research focuses on the consumer market as it aims to specifically examine end-user adoption of security measures, rather than adoption at the business or enterprise level, and is restricted by the (in)availability of the measure in specific markets.

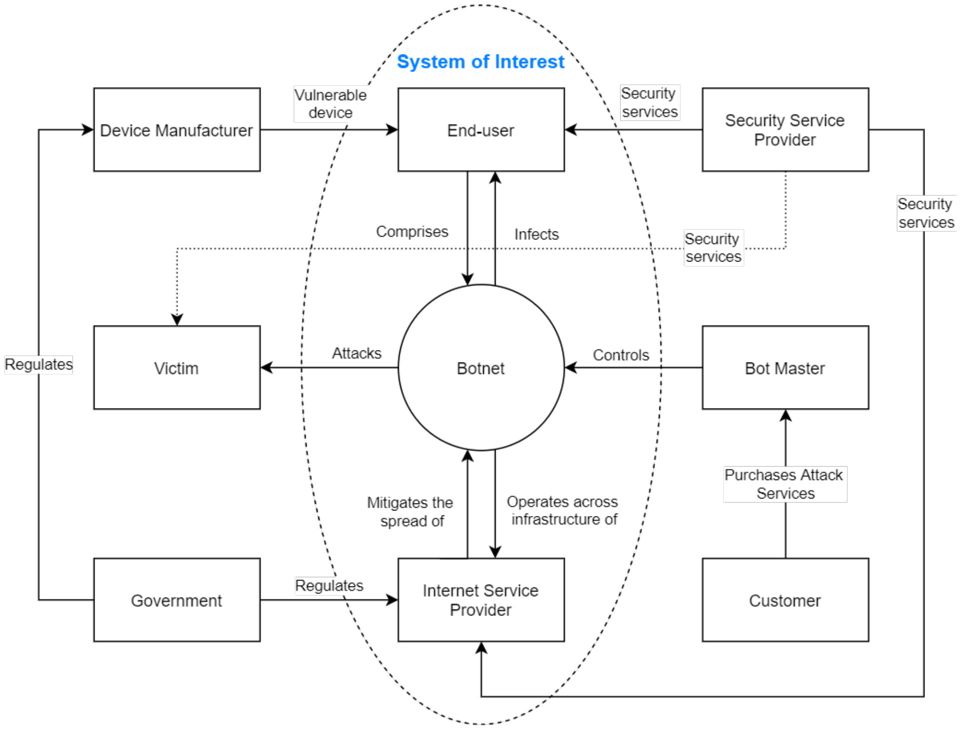


Figure 3.1: Context diagram of the system of interest.

- **Consumer market**: the consumer market is the market of interest for the intervention, as it most closely resembles the population targeted by the study. The consumer market likely contains the largest number of vulnerable devices due to the lack of structured security policies and measures imposed by corporate entities.
- **Business market**: the business market is expected to be less suitable for the experiment for multiple reasons. Primarily due to the inherent difficulties of contacting the correct people at a business, those who are responsible for the infected device(s). Moreover, as indicated by earlier research, the organizational context of the business market makes it less suitable to study the adoption of security measures by end-users (who are predominantly consumers/home-users).
- **Wholesale market**: the wholesale market constitutes service providers that use KPN's infrastructure and network to provide their services. They are therefore not the end-users which the study seeks to examine the behaviour of. Identifying and contacting end-users within the wholesale customer's domain is not possible.
- **Mobile market**: the intervention that is being studied as part of the experiment is not available for subscribers to KPN mobile services (beyond being connected to a network that is behind the filter). As such, the mobile market is disregarded as a source of subjects for the experiment.

3.4. The Intervention

The worldwide problem of internet security generates a large amount of costs for businesses and governments; phishing, DDoS, and ransomware attacks all contribute to these growing costs and the associated need to provide cost-effective security. ISPs are well positioned to mitigate malicious activity such as botnets on home users' and SME networks (Huang et al., 2007; Richards & Smith, 2007), who typically lack both adequate security and the incentives to invest in security plans that approach the socially optimal level of security and thereby resulting in a deterioration of the security of all users (Rowe et al., 2011). ISPs benefit from information asymmetry and economies of scale benefits that allow them to provide security at a lower cost, particularly to home users and SMEs.

3.4.1. Incentives and Barriers

ISP-based security services could offer new revenue sources, as well as opportunities to build customer and brand loyalty. M. Van Eeten & Bauer (2009) argue that, instead, ISPs often face disincentives in disconnecting infected machines from their networks, as this is likely to result in customers contacting the ISP's support lines and thereby imposing costs on the provider. Simultaneously, internet service providers face economic and judicial barriers in the development of convincing business models for providing security to their customers.

Technical service costs Costs associated with operating security services, particularly those which require active intervention on the end of the internet service provider. Capital and labor investments are required for tasks such as bot identification and infection remediation, while evasion techniques developed by threat actors continue to make these costs uncertain in the future.

Customer service costs Additional costs are associated with successfully notifying customers of security incidents. Emails and other communications may be regarded as spam, phishing attempts or marketing materials, while phone calls are relatively expensive and identifying the user or computer associated with a breach is similarly costly.

Legal limitations and costs Often times, contracts prevent ISPs from filtering internet traffic or performing detailed analyses. The potential implications of increased liability on the end of the ISP is an additional problem; customers might become overly reliant on the security services provided by their internet service provider and subsequently hold the ISP (partially) responsible in case of security breaches.

Nevertheless, ISPs have been increasing their efforts to fight malware (M. Van Eeten & Bauer, 2009), with the vast majority including measures such as quarantining infected machines and aiding end-users with remediation efforts, despite there often being no regulatory reason for them to do so. While the costs associated with remediation vary depending on the extensiveness of the programme, the costs associated with remediation support by ISPs alone exceeds that of phishing and scams (R. Anderson et al., 2013).

Customer support and abuse management is a key incentive for ISPs to engage in efforts to improve end-user security as customer support constitutes a significant fraction of their costs (M. Van Eeten & Bauer, 2009) that could potentially be eliminated with the use of (non remediation-based) security efforts. The authors note that while ISPs may not be formally responsible for customers' machines, they tend to be the go-to contact in case customers experience issues with internet access regardless of the nature of the issue. Incentives to provide such services for 'free' (that is; their costs are included in the service rate rather), are found in the experience that customers tend not to want to pay for such services on their own while antivirus licenses purchased by the ISP are often fixed costs. ISPs may also choose not to engage in contacting customers all together, thereby saving the direct costs associated with security efforts, but this has negative direct and indirect costs resulting from potential blocklisting of the ISP or its customers.

Brand damage and reputation effects present further incentives, as ISPs often want to present themselves as responsible businesses that provide safe services for their customers. Infrastructure expansion is another point of contemplation and incentive to engage in network security efforts as traffic growth resulting from malicious activity has outstripped the rate at which infrastructure is expanding, thus forming a problem for future network reliability if nothing is done to limit the excess traffic. Lastly, the reciprocal nature of security relations among ISPs and other security-related organisations offers a reason for them to act on cases identified within their networks.

3.4.2. The KPN Malwarefilter

According to Rowe et al., ISP-based security solutions can generally be distinguished into three categories: external, internal, and hybrid (partially internet, partially external) solutions. Fully external solutions such as providing security advice or free antivirus software. Fully internal solutions such as network-based filtering or walled-garden like measures. Partially external/internal solutions, usually policies imposed on the user that forces them to contribute to the prevention of unwanted traffic. KPN itself employs both internal and external solutions; they provide free security software and advice in the form of products like 'KPN Veilig', as well as network-level filtering solutions such as the KPN malwarefilter which is the subject of this study.

Figure 3.2 presents a simplified overview of a typical home network and the manner in which it is affected by the presence of (malicious) actors. The malwarefilter intervenes in this environment mainly by blocking outgoing DNS traffic directed at known malicious domains. In other words, it blocks requests to find the server addresses of domains which are known to be associated with malicious activity. This has both implications for the technical aspects of internet connectivity and the use of internet connected devices, as well for the user experience of customers that choose to enable the malwarefilter.

User Experience

Customers can enable the malwarefilter through their personal KPN service management environment ('MijnKPN'). Upon enabling the intervention, the customer's modem attempts to connect to a different DNS server; one that serves DNS responses in accordance with the malwarefilter's policies. Once enabled, the malwarefilter monitors DNS requests made by devices connected to the modem (wired and wireless) and returns modified responses in case a request is made for a known-malicious domain.

The feedback provided to the users of the malwarefilter are blockades observed when navigating to malicious websites in the browser. If a device located behind the KPN malwarefilter tries to do so, it will be met with a page that notifies the customer of the malicious intent of the website (the exact message that is displayed depends on the browser used by the client). The fact that feedback is limited to errors such as those displayed by browsers introduces a limitation in the degree to which customers might be aware of whether the malwarefilter is operational and to what extent it has succeeded in blocking malicious requests. Devices that require little interaction by the end-user which have been making such malicious requests may therefore not be identified as having been compromised due to a lack of visibility of the malwarefilter's functioning; it currently does not provide more extensive logging that can be accessed by its users or other overviews of legitimate and illegitimate activity.

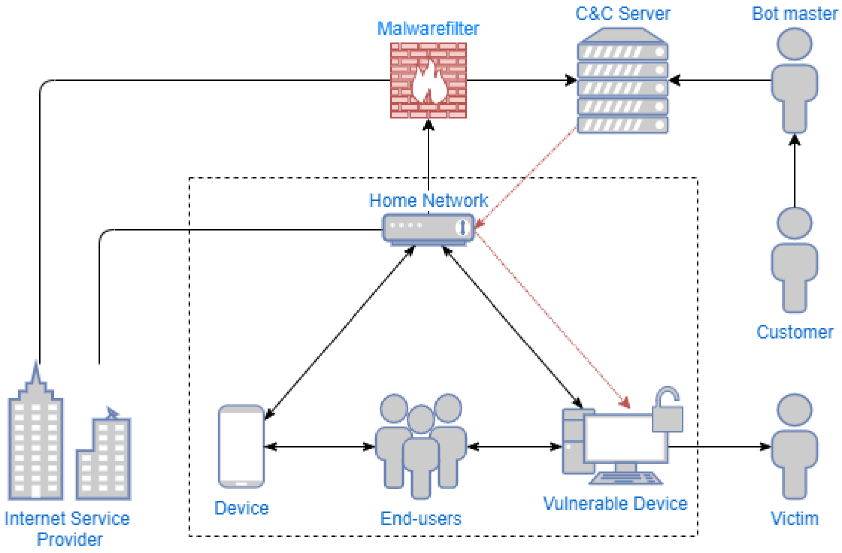


Figure 3.2: The experiment's ecosystem.

Internet Connectivity
From the point of view of malicious actors, the malwarefilter prevents devices from inadvertently downloading their malicious software from servers spreading the malware binaries (either behaviour triggered by a threat actor or through the manipulation of end-users to perform these actions). Many types of malware operate by scanning for vulnerable devices and executing remote procedure calls that force the vulnerable device to contact command and control servers and download malware binaries. By denying the vulnerable device the ability to locate the C&C server it becomes unable contact the server and download the malware (save alternative methods of establishing contact). Advanced malwares may employ more sophisticated techniques to self-propagate, reducing the mitigative strength of DNS blocklisting techniques such as those employed by the malwarefilter.

Secondly, it prevents devices from taking part in coordinated strikes against targets communicated by command and control servers. Most malware families rely on the dissemination of attack commands by C&C servers. By hindering infected devices' ability to contact these servers (by blocking DNS requests aimed at the malicious domain) coordination of such attacks can be frustrated.

As is clear from the modus operandi of the malwarefilter, the efficacy of the filter in blocking malicious activity is limited largely by its reliance on (1) having an up-to-date list of malicious domains, and (2) the necessity for malware to locate malicious domains. Advanced malwares that operate using peer-to-peer propagation methods or employ more sophisticated techniques to establish contact with. Additionally, the most state-of-the-art botnets rely less and less on communication with command and control hubs, instead delegating a significant fraction of the malicious activity required for propagation and communication to peer-to-peer techniques. Large amounts of potentially rapidly changing domains or associated servers, as is the case with Fast Flux networks, further frustrate the functioning of the malwarefilter.

3.5. Customer - ISP Relationship

There is a need to account for the relationship between the end-user and the ISP and the manner in which it affects the end-users' overall views of the ISP's services and consequently the willingness of customers to enable services such as the malwarefilter. Such a relationship may perhaps best be qualified as one of trust; a customer's perception of a service provider's ability, integrity, and benevolence (Deng et al., 2010). Customer trust and satisfaction have long been subjects of study in the context of attitudinal and behavioural loyalty, the willingness to purchase and repurchase various services from the same brand or provider (Thaichon & Quach, 2015).

Trust, in the context of cyber security, may relate most strongly to a belief that the party that provides security does so in a responsible manner (i.e. the services they provide do not negatively affect their users), and that the experience offered to the end-user is a reliable, consistent one. The degree to which an individual customer trusts the ISP may affect whether they perceive the security services managed by the ISP to be useful and capable, whether the potential costs outweigh any perceived drawbacks.

Research has generally found that factors such as security and privacy perceptions affect the trust expressed in service providers in technological environments, and indicate the existence of a reciprocal relationship between trust factors and (perceived) service quality (Chiou, 2004; Yunus et al., 2018). Thus, by ensuring the quality of services (the reliability and stability of networks, the minimization of negative effects or side-effects of their use), ISPs may improve customer loyalty towards the brand which in turn may result in improvements in loyalty behaviour that can aid the ISP in maintaining the quality and reliability of their services. The provision of centralised security measures display this same reciprocal potential, where improvements to the overall reliability of network services and reductions in negative customer experiences realised by enhanced security stand to improve loyalty behaviour and attitude towards the ISP (Thaichon et al., 2014).

4 Methodology

This chapter explicates the methods used to answer the sub questions and, consequently, the main research question. As delineated in the research plan in Chapter 1, the study aims to establish which factors motivate or demotivate end-users in the adoption of centralised security services, how effective such services are at mitigating malicious activity, and whether or not the data generated by these services may be used to provide early detection of compromised devices. In order to answer these questions, a mixed-methods approach comprising both qualitative and quantitative research methods is chosen.

Adopting a purely quantitative approach would forego the fact that cyber security has a distinctly human aspect to it. User knowledge, attentiveness, and willingness to adopt security measures are key determinants for the efficacy of security policies and controls. In the context of the presented research question, one of the deciding success factors of the security measures is the willingness of users to start using (and continue using) the mitigation service. Qualitative methods are best suited to capture such potentially complex motivations. A quantitative approach, on the other hand, facilitates the robust and methodical application of metrics to assess the impact of platform adoption on the presence of malicious activity, which is easily captured numerically and described using statistical methods.

Thus, the mixed methods approach allows the research to capture both those complex features that are not easily quantifiable, such as the factors that prove a threshold to service adoption, and those were quantitative metrics provide a robust method to assess the impact of the employed measures, such as the efficacy of the service in mitigating malicious traffic. However, associated with the benefits of a mixed methods approach are inherent disadvantages. The analysis of qualitative data is often complex and time-consuming. To alleviate this issue one might be forced to reduce the sample size, reducing the power of statistical measurements and the ability to apply conventional standards of reliability and validity, or incur a loss of accuracy or depth as a result of the quantisation of qualitative data (Driscoll et al., 2007)

Section 4.2 elaborates on the methods used to collect the required data, and the sampling methods through which the subjects and data items are selected. Section 4.3 explains how the data is managed and processed with respect to, and in compliance with, company policies and privacy and ethics guidelines. Section 4.4 goes into the methods employed to analyse the collected data and the sample sizes required to perform these analyses with sufficient rigour.

4.1. The Experiment

The host company, KPN, provides access provides to systems through which it monitors and responds to malicious activity on its networks. The department that operates these systems and which is responsible for managing incidents and contacting affected customers is the Abuse Desk. Different subsets of these customers are contacted or otherwise included in the experiment in order to passively collect data through malware infection monitoring and actively collect data through research interviews and a questionnaire.

Two sets of subjects are sourced from the Abuse feed in order to collect the necessary data. The first group of subjects is comprised of customers with an identified malware infection in the months of February and/or March of 2021, but excludes customers with a malware infection of the types of either Qsnatch or VPNFilter due to parallel research efforts at KPN. The selected customers receive a notification about their recent malware infection and are invited to enable the malwarefilter. Research interviews are performed among a subset of these customers, sampled from both those end-users that enabled the malwarefilter in response to the notification and those that did not. Malware infections are monitored for all customers that received the notification regardless of whether they enabled the malwarefilter or not, as well as for a number of individuals that did not receive the notification.

A second group of subjects is comprised of customers with an identified malware infection in the period January to June of 2021 (thus, a much larger group that encompasses the first in order to account for generally low response rates of surveys). These customers receive a notification inviting them to fill a questionnaire. This group of customers is further supplemented by distributing the questionnaire through the KPN forum, both in order to improve the number of responses and to include subjects beyond the population of customers in the Abuse feed. Figure 4.1 visualizes the sources and sizes of the various groups included in the experiment.

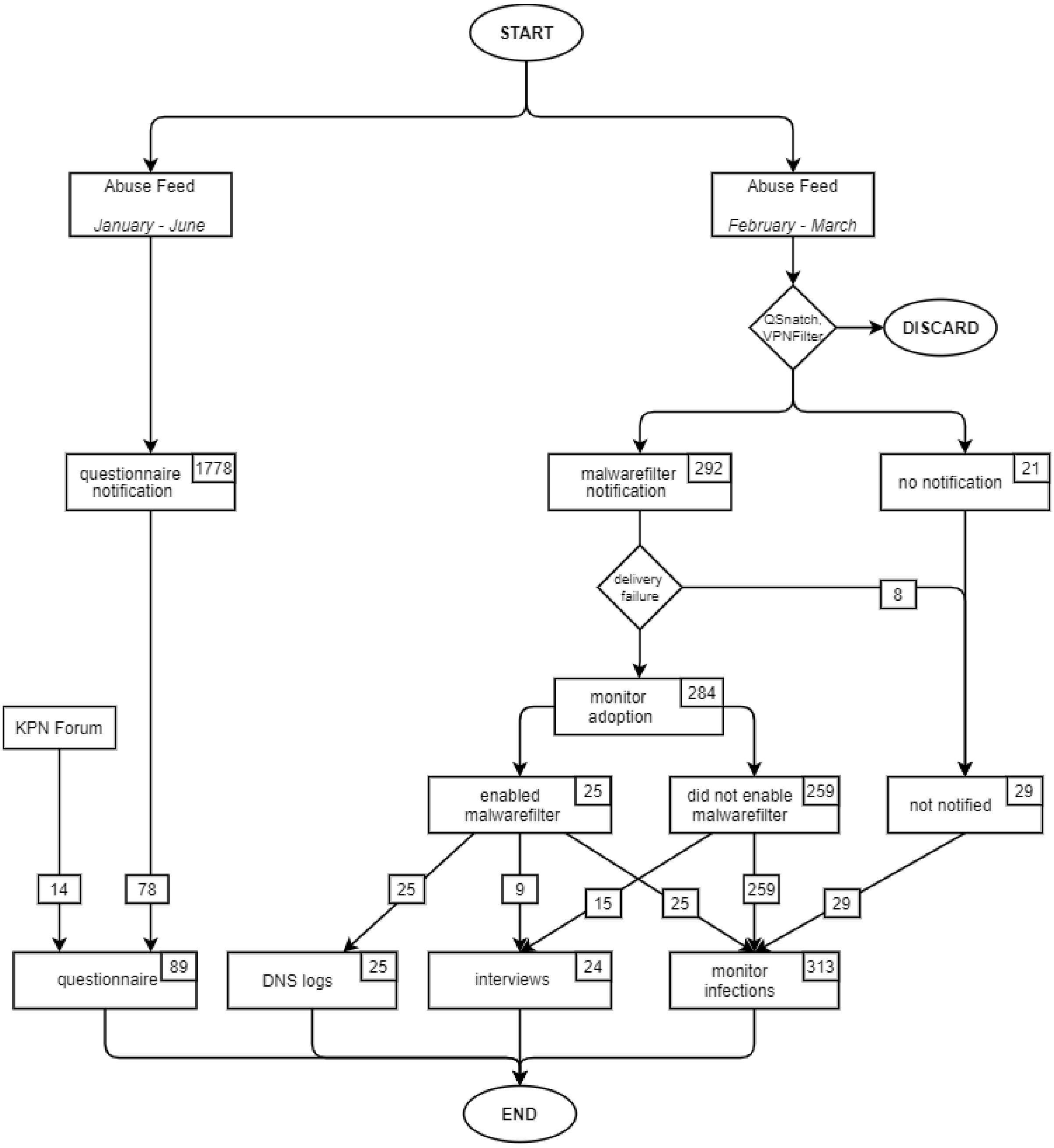


Figure 4.1: The experiment protocol.

Malwarefilter Notification

The sampled subjects receive an email notification inviting them to enable the intervention, the KPN malwarefilter. In line with the findings and improvements to remediation messages proposed by Altena (2018), the contents of the email are established around a number of guiding principles for effectively notifying customers (Table 4.1).

The notification briefly informs the subjects about the nature of their selection as a random customer selected to take part in the study, or due to a recent malware infection. The customer is invited to enable the KPN malwarefilter,

Table 4.1: Guiding principles for effective notification messages, adapted from Altena (2018)

Guideline	Suggestion	Application
Content	A clear problem statement	Short message conveying the nature of the notification in the context of a recent malware infection and/or study.
	A description of possible consequences	Information about malware, its manifestation, and the consequences of an infection.
	Actionable advice to avoid these consequences	Presenting three concise steps that can be taken to enable the malwarefilter.
Language	Short sentences and a short length of the main message	Sentences have been kept short and to the point where it does not impact the quality of information conveyed.
	An explanation in layman's terms of the problem and steps that can be taken to prevent it in the future	A brief explanation of malware, the consequences of an infection, and a link to additional information. A brief explanation of the malwarefilter, how it can protect customers' devices, and a link to additional information.
	Encouragement to take action to prevent future problems	A brief explanation of the protection provided by the malwarefilter, and a link to additional information about the service.
Layout	A logical ordering of the overall information	Main message and action steps at the top of the message. Inclusion of background information thereafter.
	Ordering of sequential elements using lists	Steps required to enable the malwarefilter outlined in a numeric list.
	Grouping of related information using headers	Background information about the KPN ID, malware, malwarefilter, and the study, is grouped and presented accordingly.
Trust	The ability to verify the authenticity of the message	Statement of the scientific nature of the notification. Inclusion of contact details of the KPN AbuseDesk and the researcher.

communicating its ability to protect them and their devices free of charge. Subsequently it presents a sequence of instructions that allow the customer to enable the service through their KPN account (their KPN ID). This constitutes the main message of the notification. The English and Dutch translations of this notification can be found in Appendix B After the main message, additional information is included on a number of subjects related to (enabling) the malwarefilter and the research effort of which the notification is a part:

- Information about the KPN ID and a link to instructions to set up or retrieve the KPN ID.
- Information about malware, how it can manifest itself, and the consequences of a malware infection for a user and their device(s), including a link to more information about malware.
- Information about how users can protected themselves and their devices by enabling the malwarefilter service, including a link to additional information about the service.
- Information about the study, such as the fact that enabling the service helps contribute to an active research project, the fact that the customer might be contacted in order to conduct an interview about cyber security and the malwarefilter, and the ability to opt out of this interview by mentioning their unwillingness to participate during the call.

Malwarefilter Adoption

In order to perform the comparisons that form the basis for the analysis of the efficacy of the malwarefilter, a method to establish the use of the malwarefilter among the contacted customers is required. Due to the unavailability of exact data on the use of the malwarefilter service, it is not possible to definitively establish which users have enabled the service or when they have done so. Thus, an indirect method to establish adoption of the malwarefilter is devised based on the DNS activity logged by the malwarefilter.

The presence or absence of DNS query activity associated with the IP addresses of the contacted customers is determined during a specific period of time. If an inquiry for a specific IP address yields some results, then we can establish that the customer to which the IP belongs must have had the service enabled at the time of the request. However, the inverse is not necessarily true. That is; the absence of DNS requests originating from the queried IP does not prove that the customer does not have the service enabled, merely that they made no DNS requests during the given time period. Based on the assumption that the vast majority of customers actively or passively use the internet on a daily basis, an estimation using this method is considered to be accurate enough to establish which customers decided to enable the malwarefilter at or before the moment of assessment.

Infection Monitoring
The KPN AbuseDesk employs an automated system for the detection of malware infections and other malicious activity within their network, based on a combination of internal and external reporters. The historical and current records of these infection cases are the means through which the prevalence of infections among the treatment and control groups can be determined. For both the recently infected customers and those who have not (specifically) been selected for their contact with the AbuseDesk the retained case records can be scanned for the customer's IP address in order to determine whether they have previously been subject to malicious activity. Subsequently the number of cases and their characteristics before and after the customer enabled the malwarefilter can be compared, as well as the difference in infection occurrence and duration between the those who enabled the malwarefilter and those who did not.

4.2. Data Collection

The study employs a mixed-methods approach towards answering the research question. This mixed-methods approach requires qualitative data collected through research interviews, as well as quantitative data collected from the Abuse feed and through a questionnaire (Table 4.2). The subsequent sections go into the sampling methods chosen to collect the research data.

Table 4.2: Research data requirements, collection, and analysis methods.

Data Type	Collection	Analysis	Used To
Qualitative	Research interview	Thematic analysis	Identify end-users' concerns and motivations with regards to the use of centralised mitigation measures.
Quantitative	Questionnaire	Regression analysis	Assess the explanatory strength of the conceptual model for the adoption of the malwarefilter.
Quantitative	Infection logs	Statistical analysis	Assess the efficacy of the malwarefilter and the practical value of such services.
Quantitative	DNS logs	Data exploration	Establish activity patterns and other characteristics of malwarefilter users' DNS requests.

4.2.1. Sampling Methods

Taherdoost (2016) distinguishes a number of sampling methods among two general methodologies; probability sampling and non-probability sampling. Probability sampling refers to a collection of methods that seek equal probability of being selected for each item in a population. Non-probability sampling methods instead are more frequently selected in qualitative research focusing on smaller samples with the intention to study real-life phenomena. Non-probability sampling methods typically simplify the data collection process, while probability sampling methods tend to introduce less bias and are more suitable for statistical inference.

As a first step to the experiment a number of KPN customers are contacted and invited to enable the intervention, the malwarefilter. These customers are selected from the group of customers that have previously been in contact with the KPN AbuseDesk. This initial convenience sampling is performed by sending the communication to customers that appear in the KPN abuse logs (which contain data about known malware infections and other cyber threats detected within the network). These customers are selected based on the type of malicious activity and the period in which the activity occurred. The historical infection case data of these customers, and the data that is generated over the course of the experiment, form the basis for the statistical analysis to assess the efficacy of the malwarefilter.

A subset of the customers that receive the invitation to enable the intervention are sampled to partake in the research interviews, to analyze end-user motivations and concerns with regards to the malwarefilter. These customers are sampled through stratified random sampling; by selecting a randomized subset of customers from both the group of subjects that enabled the malwarefilter in response to the invitation, and those who did not enable the malwarefilter. Since it is unlikely that the contacted customers enable or refrain from enabling the intervention equally, stratified random sampling allows both of these groups to be represented in the research interviews while minimizing bias that might be introduced through other (non-probability) sampling methods.

It should be noted that both the convenience sampling method used to form the overall sample, and the stratified sampling method used to select subjects for the research interview, are subject to the voluntary nature of the research. While the case data used to assess the malwarefilter's efficacy can be extracted from the ISP's systems without the need for active participation by the subjects, the collection of data during the research interviews and the extent to which this data represents the population is limited by the willingness of the contacted customers to take part in the interviews.

4.2.2. Research Interviews

Jacob & Furgerson (2012) present a guideline for interviews in qualitative research, noting the importance of a script or protocol in order to tackle a number of issues, such as inadvertent neglect to share important information with interviewees and the explicit collection of consent. Interviews should be guided by research, and contain open-ended questions that allow interviewees to take their answers in a number of directions. Jacob & Furgerson identify the length of the interview as another point of importance, noting that it may impact both the willingness of subjects to participate and the quality of responses (especially among specific groups, such as the elderly). Based on these guidelines, a protocol for a structured interview is defined (Figure D.1).

The introductory part of the interview is focused on establishing the identity of the interviewee and the intention of the interview. The customer is inquired about whether they have received the email notification about the malwarefilter and whether they have time for an interview. If the interviewee did not receive the email, does not have time for an interview, or does not want to participate or reschedule, they are discarded from the research and a new subject is chosen for the interview. If the interviewee is willing and able to participate, informed consent is established and the interview proceeds with the core questions (Figure D.2).

The main section of the interview focuses on establishing the motivations and concerns customers have with regards to enabling the malwarefilter. Questions are posed on the use of internet connected devices, perceptions of online threats, security, and responsibility. The extent to which the interviewee employs security measures and their motivations (not) to do so are established, including motivations or barriers in the adoption of the malwarefilter specifically and perceptions of advantages or drawbacks associated with the use of such services. Lastly, the perceived role of the internet service provider as the supplier of the malwarefilter service is explored by inquiring the interviewee about their perceptions of the ISP as a provider of (security) services and any implications this may have for them. Upon conclusion of the main part of the interview, the interviewee is inquired about basic demographic information: gender, age, and level of education, and thanked for their participation.

4.2.3. Questionnaire

A questionnaire is distributed among another subset of customers identified in the Abuse feed in order to gather data that can be used to validate the model proposed in Chapter 2. Interviewees are asked to rate several statements on a five-point Likert scale (ranging from strong disagreement with the statement to strong agreement). The average scores of these statements per associated construct form the basis for regression analysis, which is used to establish the influence of these constructs on the rate at which customers indicate intention to adopt the malwarefilter. Each of the statements (three per construct) has been formulated in such a way that strong disagreement is associated with lesser influence of the associated construct. Conversely, strong agreement corresponds to a greater influence attributed to the construct associated with the statement.

The questionnaire items are largely items from earlier works on the adoption of protective measures in an information security context, adapted in such a manner to fit the context of the centralised DNS-based malware mitigation platform that is the subject of this study. The works of Tsai et al. (2016) and Martens et al. (2019) provide items on perceived vulnerability and severity respectively. The items on on perceived response efficacy and perceived self-efficacy are similarly adapted from those used by Martens et al. and Ophoff & Lakay (2018). Y. Lee & Larsen (2009) supplies the items on perceived response costs. The items on subjective norm are adapted from those presented by Yoon et al. (2012). The findings of Haag et al. (2021) are used to construct items on maladaptive coping with emotions as per the recommendations presented in their study. The full list of items comprising the instrument can be found in Appendix F.

In addition to these items that seek capture end-users' perceptions of the constructs in the conceptual model, respondents are asked a number of questions about demographic variables or other features that define them or their use of internet connected devices. This includes questions about the types of devices respondents own, whether they have previously experienced malicious cyber activity and whether they are concerned about it, their security habits and practices, age, education, and experience with the use of technology.

4.2.4. Infection Cases

The identification of malware-infected devices is an automated process based on in-house detection systems such as honeypots, and third-party solutions such as infection reports. The primary sources for infection identification are the reports published by the Shadowserver foundation (*The Shadowserver Foundation*, 2021). These reports contain the IP addresses of infected devices, alongside other information such as the autonomous system (AS) number to which the IP belongs, information about the geographical location of the IP, the type of malware that has been detected. By determining which records contain an AS number that is associated with the KPN network, and by extracting the IP address from the record, KPN can identify customers whose networks originate or are otherwise associated with malware or other malicious activity. Identified malware infections and other types of malicious activity are logged in the AbuseHQ system, from where they can be downloaded in CSV format to perform custom data analysis.

The Shadowserver reports are compiled using a combination of methods. The primary method of data collection is to perform a large-scale scan of the IPv4 internet on a daily basis. Domains that point to IP addresses associated with malware activity (for example, IPs that act as C&C servers) are integrated into the sinkhole infrastructure (which reroutes the traffic intended for these domains, thus cutting the line of communication). Additionally, data is collected through honeypots (which deliberately attract malware) and the analysis of malware samples. These reports are then sent out daily to network operators, governments, law enforcement agencies, and other interested parties.

Malicious activity identified in the KPN network and the networks of its subsidiaries is logged in the AbuseHQ system. AbuseHQ is a SaaS (security-as-a-service) abuse management platform that provides statistical and other insights into network abuse and cyber threats. The KPN Abuse Desk employs a variety of playbooks (automated warning systems and notification procedures for customers) tailored to specific cyber threats. The cases logged on this platform can be inspected and filtered for a variety of features, such as the IP address from which it originates, the type of malicious activity, the type of malware associated with the activity (if applicable), the periods within which the malicious activity has been detected, and the groups the cases are associated with (based on market segments or other distinguishing features).

4.2.5. DNS Logs

The DNS data is obtained through a PowerDNS based platform. The platform records the DNS queries made by internet connections that make their DNS requests through the malwarefilter service's dedicated DNS servers. While it is not possible to extract a complete dump of all DNS queries in a given period, the platform provides a way to download a CSV file containing the most recent queries made by a single IP address. This process can then be repeated for multiple IP addresses. At the time of the study, the platform logs the requests made during the most recent 30 hours. The DNS log records contain the following fields and their associated information as detailed in Table 4.3.

Table 4.3: Non-empty fields contained in the malwarefilter DNS logs.

Field	Description	Example
timestamp	ISO 8601 timestamp	2021-05-19-T14:00:27.470Z
unix-timestamp	Unix-timestamp	1621432827.470615
requestor	IPv4 or IPv6 address of the requestor	172.16.254.1
responder	IPv4 or IPv6 address of the responder	2001:0db8:85a3:0000:0000:8a2e:0370:7334
type	Response type	A, AAAA, TXT, SRV
question	Domain requested by the requestor.	www.google.com.
answers	Answer(s) provided to the DNS query.	123.456.789.012
policy-reason	Reason for response modification.	rpZone
rcode	Response code in number (integer) format.	0
rcode-name	Response code in a named (string) format.	NoError
latency-ms	latency of the DNS request in milliseconds.	0.43

4.3. Data Processing

An important part of any research effort, especially those undertaken today, is to ensure that data is properly gathered, stored, processed, and published. This data management process should be executed in accordance with guidelines set out for research efforts, and the policies of companies that carry a (partial) responsibility for the data. Of particular interest in light of this study are the human ethics norms and guidelines (as the study involves human subjects; the customers of KPN), as well as privacy policies, legislation, and guidelines (as the study involves sensitive data).

4.3.1. Data Management

The data will be managed (that is; collected, stored, processed, and published) in accordance with a data management plan approved by the Human Research Ethics Committee of the Technische Universiteit Delft, and in compliance with the policies and regulations at KPN. Broadly, has the following implications for the management of the required and produced data:

- Interview data is collected through, and stored on, proprietary systems of KPN. The interviews are performed by the researcher in the capacity of an employee of the KPN AbuseDesk.
- Interview transcriptions are provided by KPN to the researcher and TU Delft, and stored in a research data repository for further analysis.
- Malware infection case data is collected by the AbuseDesk and its systems in an automatic process. The case data are made available to the researcher for further analysis. The aggregated, anonymous insights of this analysis are included in a research report.
- Malwarefilter users' DNS activity is recorded and kept for a short period of time by the DNS management department of KPN as a standard component of the malwarefilter service. These logs spanning a limited time span are made available to the researcher for the purposes of data analysis. The aggregated, anonymous insights of this analysis are included in a research report.

4.3.2. Ethics and Privacy

Similar to data management, the Human Research Ethics Committee of the Technische Universiteit Delft must approve of the research effort and its compliance with ethical guidelines. The primary ethical considerations are related to collection of (personal) information from interviewees, and the use of this information in such a way as to prevent privacy infractions.

In order to collect data from the subjects in an ethical fashion, subjects must explicitly provide their consent to participating in the study. Informed consent is established at the onset of the interview, by explicitly asking the interviewee whether they understand that their participation is voluntary and anonymous, and whether they agree to a recording of the interview. After consent has been established, the interview proceeds. The information provided by the interviewee is transcribed and anonymized in a manner that is consistent with the ability to perform the required analyses and the privacy of the subjects. Interview transcriptions contain no personally identifiable information.

A secondary source of data are the infection cases and DNS log data, which are not collected by or with the intervention of the researcher, but whose use is nevertheless subject to ethical and (primarily) privacy concerns. KPN, in their role as internet service provider, have the legal grounds to collect data on their customers in manners consistent with (international) law, which allows them to monitor and contact customers for purposes of abuse remediation. The position of the researcher as a research intern (i.e. an employee) at KPN provides the opportunity to use and present the anonymized, aggregated insights obtained from the malware infection cases in a research report. Infection case data does not leave the premises of the KPN systems.

As users enable the malwarefilter service, they consent to the collection and analysis of DNS traffic. Activity logs spanning a short time period (approximately one week) are maintained for troubleshooting purposes. These logs are provided to the researcher - as a research intern at KPN - for analysis. Similar to the results of the other analyses; solely the aggregated and anonymized insights obtained from these analyses are included in a research report. DNS log data does not leave the premises of the KPN systems.

4.4. Data Analysis

A number of methods are used to analyse the data gathered through the interviews and the infection monitoring systems of KPN. First and foremost, the interview data is analyzed using thematic analysis in order to establish themes and patterns in the subjects' answers. The initial code for this analysis is informed by the literature review of Chapter 2. Secondly, the analyzed interview data are combined with the monitoring results and described using statistical methods.

The ATLAS.ti software package is used to perform qualitative analysis of research interview data. IBM's SPSS and the Python programming language (primarily; the data analysis and manipulation tool Pandas (McKinney, 2011)) are used to perform quantitative analysis. Data visualization is conducted through the visualization library MatPlotLib (Hunter, 2007).

4.4.1. Thematic Analysis

The qualitative data collected through the research interviews is analyzed using the thematic analysis method. Thematic analysis is a widely used qualitative analytic method within a variety of fields, notably in the field of psychology, for the identification, analysis, and reporting of patterns within data. Thematic analysis has a number of advantages relative to other qualitative analytic techniques, such as flexibility, accessibility (of the method, as well as the results), and the ability to generate unanticipated insights (Braun & Clarke, 2006).

The process of conducting thematic analysis can be roughly captured by six steps, as delineated in Table 4.4. After the dataset has been constructed, the researcher(s) should familiarize themselves with the data in order to obtain ideas for an initial set of codes. This codebook is subsequently expanded and improved based on the iterative coding of an increasingly larger fraction of the data. Related codes are combined into categories or themes which are similarly refined over multiple iterations. Based on the sufficiently refined codebook/themes a story is developed, the findings are reported and used to answer the research question(s).

One of the key decisions to be made in applying thematic analysis is whether to perform inductive or deductive identification of the codes and themes, or whether to apply a mix thereof (Braun & Clarke, 2006; Gale et al., 2013). An inductive approach toward code book generation aims to identify recurrent features from the gathered data. A deductive approach instead generates the (initial) codes and themes according to existing theories and frameworks. A commonly used hybrid approach generates the initial code book based on theory, and refines the themes and codes as the analysis progresses.

Step	Action	Description
1	Familiarization with the data	Transcribing the recordings and noting interesting features and initial ideas.
2	Generating initial codes	Defining a set of codes based on a systematic inspection of (part of) the dataset.
3	Searching for themes	Collecting related codes into themes and gathering relevant data for each theme.
4	Reviewing themes	Iterative analysis and refinement of the code and composite themes.
5	Defining and naming themes	Refinement of the specifics and overall story; generating clearly defined themes.
6	Reporting the findings	Relating the analysis to the research question, producing a report of the findings.

Table 4.4: The six steps of thematic analysis, adapted from Braun & Clarke (2006).

Researchers conducting thematic analysis should be wary of mistakes in applying such a flexible and easy-to-use method. Braun & Clarke note several pitfalls in the application of thematic analysis. One might fail to actually perform any analysis at all, instead producing a collection of extracts with little analytical narrative. Associated with this pitfall is the use of the data collection method as the reported themes, in which case it becomes difficult to make sense of the pattern of responses. A third pitfall is an unconvincing analysis where the themes show little consistency or too much overlap, thereby failing to capture the majority of the data. A fourth and fifth pitfall relate to the potential mismatch between the data and analytical results, and a mismatch between theory and the analytical claims respectively. These pitfalls can be addressed through a solid foundation of the analytical groundwork in theory and the thorough documentation of assumptions and procedures.

Inconsistencies in the interpretation of data, for example because of interpreter bias, are mitigated by refining the code based on consensus between multiple coders using intra- and inter-observer agreement metrics. Well-known metrics for agreement between two coders (inter-rater reliability) are Cohen's kappa and its more-than-two coder equivalent Fleiss' Kappa. Qualitative analysis software such as Atlas.ti employ variations of these metrics such as Krippendorff's Kappa, which nonetheless operate largely on the same principle. These kappa metrics measure agreement between raters whom classify items into a number of mutually exclusive categories. Cohen's kappa is defined as:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

p_o denotes the relative observed agreement between raters. p_e denotes the probability of agreement based on chance. Thus, if two raters agree completely Cohen's kappa equals 1; if the agreement does not exceed what would be expected by chance then Cohen's kappa equals 0. Negative kappa values imply that the agreement between raters is worse than would be expected according to a random assignment of ratings. A range of values have been defined that give some indication about the level of agreement based on the achieved kappa value. This segmentation of the kappa values and their interpretations can be found in Table 4.5. In literature and practical applications, a Kappa value of 0.70 or greater is generally considered to indicate good inter-rater reliability, although lower thresholds may be acceptable for more exploratory research.

Table 4.5: Interpretation of kappa values and agreement levels.

Kappa	Agreement
≤0	Poor
0.01 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	(near) Perfect

4.4.2. Statistical Analysis

Descriptive statistics are used to describe the dataset obtained from the interviews and the observations made through the infection monitoring systems of KPN. These statistics describe, for example, the distribution of demographic variables such as age and gender among the interviewees, or the distribution of infections observed in the target population and the test and control groups over time. This is typically achieved through a combination of numeric descriptors such as the mean and standard deviation of a variable, and the visualization of the distribution of one variable (univariate analysis) or the relationships between multiple variables (multivariate analysis). The descriptive statistics are meant to provide an overview of the acquired dataset.

Subsequently, inferential statistics are used to describe the relationship between experiment groups and the monitoring observations, and yield insights into the reliability of conclusions drawn from observations within the experiment. Two main methods of inferential statistics are employed as part of this study; hypothesis testing and regression analysis.

Hypothesis Testing

The efficacy of the malwarefilter can be measured in a variety of ways. Given the limitations imposed by the environment of the study and the availability of resources, the primary metrics for efficacy rely on the case information generated by KPN's incident management system. Two types of metrics can be distinguished; within-group metrics and between-group metrics. Within-group metrics can be used to compare the situations before and after the introduction of the intervention (that is; the enabling of the malwarefilter by the contacted customers). Between-group metrics can be used to compare the situations of the different groups, primarily the difference between those customers that enable or use the malwarefilter service and those that do not.

- The number of infections that occur in the control group compared to the number of infections in the treatment group.
- The number of infections that occurred before the treatment compared to the number of infections that occurred after.
- The duration of infections that occur in the control group compared to the duration of infections that occur in the treatment group.
- The duration of infections that occurred before the treatment compared to the duration of infections that occur after the treatment.

Statistical hypothesis tests such as the parametric T-test and non-parametric Mann-Whitney-U test can be used to compare samples from two groups in order to determine whether there are significant differences in the distributions underlying some features of these groups. Such tests can thus be used to assess whether an identified effect is likely the result of chance or whether the observed differences are significant enough that they are likely to be the result of something other than chance (e.g. the introduction of an intervention such as the malwarefilter in one of the groups). Table 4.6 presents a number of hypothesis tests that are performed to assess both the efficacy of the malwarefilter as well as account for other effects that might affect the results.

Comparison tests such as the parametric T-test and non-parametric Mann-Whitney U test can thus be used in case of two samples (comparing treatment and control groups). The comparison of more than two samples or groups requires techniques such as the parametric MANOVA (multivariate analysis of variance) test, or the non-parametric Kruskal-Wallis test.

Regression Analysis

Regression analysis is used to assess the extent to which the constructs defined in the conceptual framework predict or affect the adoption of the malwarefilter. The two most common types of regression analysis methods

Table 4.6: Hypotheses about the number of infection cases and their duration.

	Null Hypothesis (H_0)	**Alternative Hypothesis (H_1)**
H1a	There is no significant difference in the distribution of malware infections across users and non-users of the malwarefilter before the intervention.	There is a significant difference in the distribution of malware infections across users and non-users of the malwarefilter before the intervention.
H1b	There is no significant difference in the distribution of malware infections across users and non-users of the malwarefilter after the intervention.	There is a significant difference in the distribution of malware infections across users and non-users of the malwarefilter after the intervention.
H2a	There is no significant difference in the distribution of infection durations across users and non-users of the malwarefilter before the intervention.	There is a significant difference in the distribution of infection durations across users and non-users of the malwarefilter before the intervention.
H2b	There is no significant difference in the distribution of infection durations across users and non-users of the malwarefilter after the intervention.	There is a significant difference in the distribution of malware infection durations across users and non-users of the malwarefilter after the intervention.
H3a	There is no significant difference in the distribution of malware infections among users of the malwarefilter before and after the intervention.	There is a significant difference in the distrubtion of malware infections among users of the malwarefilter before and after the intervention.
H3b	There is no significant difference in the distribution of infection durations among users of the malwarefilter before and after the intervention.	There is a significant difference in the distribution of infection durations among users of the malwarefilter before and after the intervention.
H4a	There is no significant difference in the distribution of malware infections among non-users of the malwarefilter before and after the intervention.	There is a significant difference in the distribution of malware infections among non-users of the malwarefilter before and after the intervention.
H4b	There is no significant difference in the distribution of infection durations among non-users of the malwarefilter before and after the intervention.	There is a significant difference in the distribution of infection durations among non-users of the malwarefilter before and after the intervention.

are linear regression and logistic regression, which operate under different assumptions about the dependent and independent variables.

Logistic regression (like other regression methods) aims to establish the equation that best predicts the value of a dependent variable Y based on a predicting variable X. Multiple logistic regression extends this model by allowing for the simultaneous estimation of the relationships between multiple independent variables X on a single binomial dependent variable Y. Multiple logistic regression approximates the following logical and mathematical relationships:

$$\{x_1, x_2, ..., x_n\} \longrightarrow y \qquad\qquad ln(y/(1-y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_3 x_3$$

Unlike linear and most other non-linear regression models, logistic regression can operate with categorical dependent and independent variables, such as a binary variable that describes whether a customer has enabled a service or whether they did not. It is therefore the most suitable method for the analysis conducted in this study, as it adheres most closely to the nature of the examined variables.

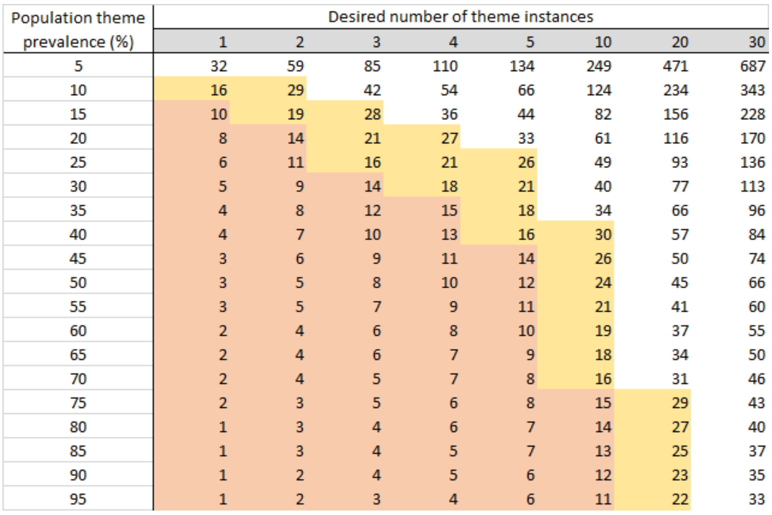
4.4.3. Sample Sizes

A number of factors strongly influence the ability of the study to assess with reasonable reliability the motivations for end-user security behaviour and the efficacy of the malwarefilter service. The reliability of conclusions drawn from the thematic analysis is perhaps most difficult to asses, although a range can be estimated based on factors such as the number of themes the study aims to identify and the desired prevalence level of these teams in the data. Estimating the sample size required to achieve sufficient reliability when performing regression analysis relies heavily on heuristics, while samples sizes for hypothesis tests are perhaps easiest to determine with great accuracy.

Thematic Analysis

Sample sizes are a topic of discussion among practitioners of qualitative research methods, with typical estimates ranging from ten to more than a hundred samples. While little consensus exists on how to determine the level of saturation that has - or can be - achieved, and even discussion on whether this is a useful metric at all (Braun & Clarke, 2006), there is agreement that the required number of samples depends both on the context of the study and the level of detail required by the study. Fugard & Potts (2015) provide one of the most widely applied tools for ex-ante sample size estimation ex-ante, based on the desired number of themes and theme prevalence within the population.

Based on an assumption of 80% power (that is; an 80% chance to identify an instance of a theme in a sample) and the tool provided by Fugard & Potts, a sample size of 15 is considered to be sufficiently capable of capturing a fair number of the most prevalent themes, as well as a number of less prevalent ones (Figure 4.2). Larger sample sizes such as 30 primarily allow for a greater distinction between the prevalence of themes, but one can expect a diminishing return with regards to new information obtained. Sample sizes greater than 30 offer strong diminishing returns, and one might expect the vast majority of themes to have been covered by the previous data points. As such, an interview sample of size 15 to 30 is sufficiently capture the required data.



Population theme prevalence (%)	Desired number of theme instances							
	1	2	3	4	5	10	20	30
5	32	59	85	110	134	249	471	687
10	16	29	42	54	66	124	234	343
15	10	19	28	36	44	82	156	228
20	8	14	21	27	33	61	116	170
25	6	11	16	21	26	49	93	136
30	5	9	14	18	21	40	77	113
35	4	8	12	15	18	34	66	96
40	4	7	10	13	16	30	57	84
45	3	6	9	11	14	26	50	74
50	3	5	8	10	12	24	45	66
55	3	5	7	9	11	21	41	60
60	2	4	6	8	10	19	37	55
65	2	4	6	7	9	18	34	50
70	2	4	5	7	8	16	31	46
75	2	3	5	6	8	15	29	43
80	1	3	4	6	7	14	27	40
85	1	3	4	5	7	13	25	37
90	1	2	4	5	6	12	23	35
95	1	2	3	4	6	11	22	33

Figure 4.2: Thematic analysis sample size tool, adapted from Fugard & Potts (2015).

An ex-post or in-medio-res analysis of the sample's adequacy is typically performed by calculating the achieved saturation at various points throughout the data analysis. Effectively, saturation measures the degree to which the inclusion of additional data items yields new information. Although the usefulness of data saturation as a concept is debated (Braun & Clarke, 2021), it remains to be a commonly used metric among qualitative analytic approaches. Saturation is thus typically used to define a point where capturing additional themes requires a disproportionate collection effort.

Guest et al. (2020) define a quantitative method to measure data saturation. By calculating the number of new themes obtained from the coding of a number of subsequent data items (denoted by the run length), and comparing it to the number of themes in an initial set of data items (denoted by the base length), the percentage of 'new information' (i.e. not previously encountered themes) can be calculated. Figures 4.3 and 4.4 visualize the metric calculation and present the parameters proposed by Guest et al. respectively. If the percentage of new information is below some established threshold, saturation confidence is sufficient for the study and no more research interviews are conducted. For this study, a base length of six, alongside a run length of three, and an information threshold of 0% are parameters chosen to calculate data saturation, representing a relatively high level of saturation confidence.

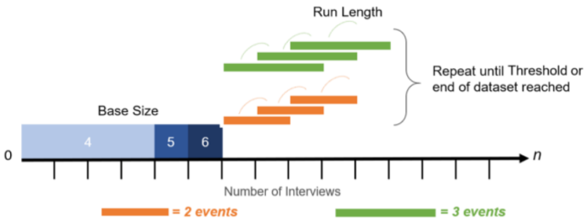
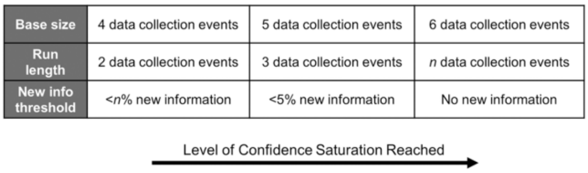


Figure 4.3: Base size and run length (Guest et al. (2020)).



Base size	4 data collection events	5 data collection events	6 data collection events
Run length	2 data collection events	3 data collection events	*n* data collection events
New info threshold	<*n*% new information	<5% new information	No new information

Level of Confidence Saturation Reached

Figure 4.4: Saturation confidence (Guest et al. (2020))

Regression Analysis

Regression analysis is used to estimate the impact of the independent predictor variables on the dependent variable of intention to adopt the malwarefilter. Although no consensus exists on the sample size required to perform a reliable regression analysis, a typical estimate is to obtain ten to twenty times the number of observations as there are constructs in the model. Thus, to validate the constructs defined in the model and their effects on the intention to adopt the malwarefilter requires a sample size between 70 and 140 according to this heuristic.

Based on estimates provided by the KPN Market Intelligence department and a pilot questionnaire distributed among 52 KPN Abuse customers, a response rate of 8 to 10 per cent is realistically expected. Based on this indication, the questionnaire must be distributed among 700 to 1750 individuals in order to obtain sufficient responses (assuming 10 responses per construct at a 10% response rate as a minimum estimate, and 20 responses per construct at an 8% response rate as a maximum estimate).

Statistical Analysis

The estimated effect size and the parameters that regulate the probability of incurring a type I or type II error are the predominant factors in estimating a sample size for reliable hypothesis tests. A type I error occurs when the null-hypothesis is unjustly rejected (thus, it constitutes a false positive; detecting an effect that is not present) and is mitigated by reducing the α parameter. A type II error occurs when a false null-hypothesis is unjustly accepted (thus, it constitutes a false negative; not detecting an effect that is present) and is mitigated by reducing the β parameter. The β parameter is often omitted in favour of 'power', which is defined as $1 - \beta$ and entails the probability of finding an effect if it is indeed present. Power analysis methods can be used to estimate the sample size required to achieve sufficient reliability of hypothesis tests.

Table 4.7 defines the parameters used to estimate the sample size for two-group independent samples hypotheses. Calculating the required sample size using these parameters yields a total sample size of 102, 51 subjects per group, in the case of the t-test or a total of 106, 53 subjects per group, in case of the non-parametric Mann-Whitney-U test. Malwarefilter adoption rates as low as 10%, thus implying an allocation ratio of 9, can be accounted under for by including approximately 300 individuals in the experiment.

Table 4.7: Hypothesis test parameters used to estimate the required sample size.

Parameter	Explanation	Value	Justification
Tail(s)	One- or two-tailed; whether to consider effects in the opposite direction.	One	The effect has an expected direction (e.g. fewer infections among malwarefilter subscribers).
d	Effect size; expected magnitude of the effect. Larger effects are easier to pick up.	0.5	Medium effect size, suggested by literature in case little is known about the expected effect size.
α err	Alpha; the probability of incurring a type I error.	0.05	Standard value for statistical tests; relatively minimal probability of picking up effects that are not real.
1 - β err	Power; the probability of not incurring a type II error.	0.80	Suggested by literature and earlier studies (Altena, 2018).
N2/N1	Allocation ratio; the ratio of subjects in each of the experiment groups.	1	Actual value depends on malwarefilter adoption rate.

4.5. Pilots

A number of pilot interviews were performed to gauge response rates and assess whether the interview questions were both comprehensible and adequate to capture the required data. Over the course of these pilot interviews, a number of iterative changes were made to the questions and related content included in the research interviews. These changes have been made in part to the order and manner in which certain questions are posed during the interview in order to maintain a more natural flow of conversation while keeping in line with the systematic nature of the interview. Appendices C and D contain the interview protocol used during the pilot interviews and the refined protocol respectively.

While the initial setup of the experiment was based on the collection of the survey data as part of the research interview, limitations in the response rates and boundaries imposed upon the duration of the experiment have resulted in the choice to instead distribute the survey in the form of an online questionnaire. An exchange with the market intelligence department of KPN confirmed that while response rates are typically considered low amongst both telephone and online surveys, they had noticed a deterioration in telephone survey response rates and refrain from performing this type of research as a result. Appendices E and F contain the questionnaire distributed to a subset of approximately 50 customers as a pilot, and the refined final version respectively. Analyses of the research interviews and responses to the questionnaire are presented in Chapter 5.

5
Analyses

Section 5.1 presents the results of the inquiries into the motivations for (non)adoption of the intervention. It presents the the results of the research interviews, analysed to identify the motivations and concerns underlying end-users' willingness to adopt the malwarefilter. Section 5.2 quantitatively validates these findings and the degree to which the conceptual model proposed in Chapter 2 is capable of predicting the intention to use centralised malware mitigation services. Section 5.3 presents an analysis of the efficacy of the malwarefilter; a comparison of infection occurrence and duration between the periods preceding and following the introduction of the intervention, and between the experiment groups. Section 5.4 presents an exploration of the DNS logs, in an effort to achieve greater insight into the activity patterns of malwarefilter users.

5.1. Security Perception and Motivation

As part of the experiment, 292 customers that experienced a malware infection in February or March of 2021 received an email communication about the malwarefilter. Eight emails could not be delivered, resulting in the exclusion of these customers from the experiment and leaving a total of 284 subjects. Of these 284 end-users, 25 had enabled the malwarefilter at the point of assessment in April, while the remaining 259 had not. Interviews were conducted among both groups of end-users until saturation was considered to have been reached at 20 data items (by which point a total of 24 items had been collected). Nine interviews were successfully conducted among the 25 customers that had enabled the malwarefilter, while 15 could be conducted among 100 randomly sampled individuals that had not enabled the service.

During these interviews, the customers were inquired about their beliefs about various aspects of the security of internet connected devices and online threats to such devices, as well as their reasons for (not) enabling the malwarefilter (and their experiences with it if applicable). Lastly, customers were inquired about their perception of views on the ISP as a provider of security services.

The data obtained through these interviews was analysed using the thematic analysis method until sufficient confidence in the saturation of the data was achieved. Table 5.1 presents the saturation confidence table in accordance with the metric suggested by Guest et al. (2020). Saturation was calculated using a base length of six and a run length of three. Saturation confidence was considered to be sufficiently high after 20 interviews, with no new information being obtained from the individual data items eighteen, nineteen, and twenty. Items 21 through 24 had already been collected and are therefore included in the analysis, but yielded equally little new information. Table 5.3 presents an overview of the demographic makeup of the group of interviewees, and their use of various device types, security measures, and the malwarefilter itself.

Table 5.1: Saturation confidence based on the occurrence of new codes in individual data items (NC item), in each run (NC run), and the percentage of new information (NI %) in the run.

	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NC (item)	2	0	1	3	2	0	1	0	1	0	0	2	0	0	0	0	0	0	0
NC (run)	26			4	6	5	1	3	2	1	1	2	2	2	0	0	0	0	0
NI (%)				15	20	14	7	2	4	2	2	4	4	4	**0**	**0**	**0**	**0**	**0**

Reliability

An initial codebook was developed in a hybrid fashion, deductively based on the conceptual model and inductively based on an initial coding of the pilot interviews. This initial codebook is subsequently refined inductively through the creation of new codes as new interview transcriptions are processed, and the merging of overlapping codes. Inter-coder reliability was assessed by a second coder coding a subset of the interview items using the refined codebook developed by the primary researcher after coding all data items. Two widely-used inter-coder reliability metrics were calculated using Atlast.ti built-in functionality; Holsti's index and Krippendorff's Alpha. Table 5.2 presents an overview of the inter-rater reliability for each of the items coded by both the first and the second coder.

Holsti's index is a percentage-agreement measure, which defines the degree to which two or more coders agree on their codings but does not account for agreement by chance. A Holsti's index value of 78.6% was achieved across the coded items, with minimum and maximum agreement values of 62.1% and 93.1% among the data items. Krippendorff's Alpha, like Cronbach's Kappa and other more intricate merics, takes into account the degree of agreement attained by chance. A Krippendorff's Alpha value of $\alpha = 0.719$ was attained across data items with minimum and maximum values of 0.338 and 0.841.

Table 5.2: Inter-rater reliability values per data item.

Data Item	Krippendorff's Alpha	Holsti's Index
ALL	0.719	78.6%
1	0.338	74.4%
2	0.808	81.7%
3	0.712	76.4%
4	0.841	93.1%
5	0.829	62.1%
6	0.641	82.0%

5.1.1. Weak passwords and suspicious links: the role and responsibility of end-users

The most prominent theme, which was almost universally identified among data collection events was the responsibility of the (primary) user or owner of an internet connected device in ensuring they are adequately secure against online threats. The overwhelming majority of interviewees mentioned that the end-user plays a key role in the security chain in one or both of the following ways: by employing good security practices and/or by actively avoiding malicious content.

The length and complexity of passwords, or the use and integrity of alternative authorization methods such as pin codes or two-factor authentication, is consistently mentioned as an important user-related aspect of device security. The use of (some degree of) security software or other services is widespread, and considered by most to be a critical responsibility of the user in securing their devices. In addition to instating strong passwords and employing software-based security, some end-users also consider the practices of regularly installing security updates, regularly changing passwords, or using different passwords for different services as an important security habit.

> "The user themselves, if you set up a weak password that you use on multiple systems then it's a matter of time until they know how to find you. Every account you register is an entry into your other accounts."

> "In the end, I myself am responsible for my pin codes for example. In the end you can be held responsible for the use of two-factor authentication and what not."

In addition to these security practices, end-users consider themselves responsible for (and generally; capable of) actively making efforts to avoid malicious content or malicious activities. This involves obtaining sufficient knowledge about, or trust in, the legitimacy of software that the user chooses to install, or the links they click.

> "In the end, the user itself is responsible for the security of their devices. Naturally, that responsibility starts with not simply navigating to random websites that serve malicious content."

5.1.2. Supporting end-user security practices: the role of suppliers and the ISP

End-users consider the role of suppliers such as device manufacturers and the internet service provider to be primarily supportive or enabling in nature. Device manufacturers have two roles in ensuring the security of their internet connected devices as identified by end-users. First and foremost, they should ensure an basic level of

Table 5.3: Demographics, device ownership, and security measure use of the subjects that participated in the research interviews.

#	Gender	Age	Education	Computer	Mobile	Smart	Security Software	Malwarefilter
1	Female	51	Higher Education	x	x	x	Yes	**Yes**
2	Male	60	Higher Education			x	Yes	**Yes**
3	Female	44	Vocational Education	x	x	x	Yes	**Yes**
4	Male	41	Higher Education		x		Yes	No
5	Male	50	Higher Education	x	x	x	Yes	No
6	Male	48	Vocational Education	x	x		Yes	**Yes**
7	Male	29	Vocational Education	x	x		Yes	No
8	Female	52	Vocational Education		x		Yes	No
9	Male	50	Higher Education	x	x	x	Yes	**Yes**
10	Male	42	Higher Education		x		Yes	**Yes**
11	Female	32	Higher Education	x	x		Yes	No
12	Male	46	Vocational Education	x	x		Yes	No
13	Male	58	Higher Education		x		Yes	No
14	Male	40	Higher Education		x		Yes	No
15	Male	56	Higher Education	x	x	x	Yes	**Yes**
16	Male	50	Higher Education	x	x		Yes	**Yes**
17	Male	57	Higher Education	x	x	x	Yes	No
18	Male	58	Vocational Education	x	x		Yes	No
19	Male	48	Higher Education		x		Yes	No
20	Female	49	Secondary Education		x		Yes	**Yes**
21	Male	24	Higher Education		x		Yes	No
22	Male	60	Higher Education	x	x		Yes	No
23	Male	43	Higher Education	x	x	x	**No**	No
24	Male	30	Higher Education	x	x	x	Yes	No

security, for example by imposing good security practices on the end-user such as requiring them to set a strong password.

> “In the first instance, the manufacturer is responsible for extraditing these devices in such a way that there must always be a strong password.”

Secondly, they should support the end-user in maintaining the security and integrity of their devices by rolling out security updates and patching known vulnerabilities.

> “In addition, I think there is a responsibility for the device manufacturer to fix security vulnerabilities that have been reported to them.’

While end-users think that device manufacturers should support the security of their devices, they typically do not consider this to be one of the manufacturers’ primary responsibilities. Reasons being that the user of a device carries the final responsibility for its security, and that device manufacturers may not have the necessary skills to ensure protection in some cases, or that ensuring the adequate security of their devices is not one of the priorities of manufacturers in other cases. Much like the supportive role of manufacturers, a significant fraction of customers, 7 out of 24 interviewees, underwrite the idea that their internet service provider is - or should be - partially responsible for ensuring a degree of protection against online threats.

> “The end-user has the first responsibility, but it would be very much appreciated if the internet service provider could have some impact there as well.”

> “I don’t think [the internet provider] is responsible, I don’t think it’s a ‘must’, but it would be nice if they could offer the customer a little something to give them some extra protection.”

5.1.3. Experience and media attention: vulnerability to online threats

In assessing the threats to online devices, familiarity with online threats play an important role. Both previous first-hand experiences with security incidents, as well as second-hand accounts such as media coverage of security incidents are identified as influencing the degree to which an individual considers themselves capable of assessing the dangers posed by online threats. 9 of the interviewed individuals explicitly mention that familiarity with security incidents, either through personal or professional experience or the coverage of security incidents by the media, plays a role in threat assessment.

> “It’s very well known and covered in the media that a lot of devices are sold with standard passwords or no passwords instated.”

Those individuals that did not mention any previous experiences with security incidents noted difficulty in assessing the potential impact of threats and the security of their internet connected devices (8 out of the 15), often due to this lack of (personal) experience or a lack of knowledge about the information technology landscape. In some cases interviewees noted that they consider themselves unlikely to be the target of malicious activity (2 cases), or that they would not be possible to prevent such online threats in the first place (3 cases).

> “I’ve never experienced negative effects [from using internet connected devices]. I do not want to commit to saying it is not possible, but I don’t think it is very likely.”

5.1.4. My device and my data: the consequences of malicious activity

End-users typically consider malicious activity on or enabled by their internet connected devices as primarily posing a danger to themselves and their interests in their daily lives; 9 out of 24 interviewees mentioned that abuse of internet connected devices might harm their personal interests. Data integrity loss on various levels is most often mentioned as a possible consequences of malware or other compromises of online devices. Online threats are sometimes considered to have negative consequences for the devices themselves (5 interviewees expressed this view), although they are also considered to pose little real, tangible threat to the functionality of a device. While it may hamper their performance to some degree, limiting device performance to such a degree that they become unproductive is not a widely held belief.

> “I don’t think it is simply a matter of these devices being infected and that they stop working at all. So I am not very afraid of that, of malware infections or whatever.”

> “I think the most troublesome consequence would be that they could perhaps access your banking. I am not very concerned about the rest; there are no other outrageously interesting things on there, but I do worry about what it would mean for my banking.”

The danger posed by a compromised device to the security and integrity of other parties (parties other than the device’s owner or primary users) is rarely recognized; only 2 interviewees mentioned possible consequences for others as a potential result of abuse of internet connected devices. The individuals that did recognize this danger placed it in the context of previous experiences or notifications of such consequences, indicating that - much like threat assessment in general - familiarity with various types of threats online plays an important role in identifying its ability to impact individuals beyond the user themselves.

> “One possible consequence of an infection could be that the device is used for other purposes such as spreading malware or other viruses. I think that was one of the indications with my system, that it was being used to spread malware.”

5.1.5. Perceptions of device security and the influence of authority on security.

The perceived security of internet connected devices regularly depends on characteristics such as device type or operating system, its age, or the extent to which the user is able to install or keep the device’s security mechanisms up to date through security updates. The ability to connect certain (mobile) devices to other networks, of which the end-user often cannot assess their security at all, is also experienced as a limiting factor in their ability to be truly secure.

> “No I don’t think such devices are generally secure against being abused by a malicious party. Of course they claim that they are, but I have my doubts about that.”

> “I have some of these smart light switches to turn your lights on and off; I have my doubts about those because there is no way to update them.”

Nevertheless, the majority of interviewees (13) expressed doubts about the security of their online devices in general. Three customers expressed that their considered their devices to be secure, but only because of the security measures they had set up themselves. Doubts in the security of internet connected devices are often rooted in a lack of knowledge about information technology or the information security landscape, making it difficult to assess the degree to which such devices are secure out-of-the-box or are adequately secured through the user’s own measures:

> “I don’t think my devices are secure enough, no, because I’m not very familiar with it. I do have one of these anti-malware applications that I check with some regularity, but it more or less stops there.”

Nearly half the interviewed subjects reported that they either enabled the malwarefilter in response to communications sent by the internet service provider as part of the experiment or before that or as part of earlier communications with the ISP or the Abuse Desk. The prevalence of this factor in spurring customers to enable and in other cases investigate the use of the recommended service - the malwarefilter - is a striking display of the impact of direct and indirect effects of authoritative parties on the adoption of security measures.

> “I did receive similar communications from you earlier, so I installed an anti-virus application.”

> “Yes I did enable the malwarefilter the other day, on the advice of the ISP itself. I received an email that I should enable it because of a potential infiltration or something like that.”

5.1.6. Trust and Privacy: the barriers of centralised security

Adopting centralised security measures implies entrusting the provider of measure with the end-user’s personal data and what this implies for the privacy of the user, as well as other parties that may obtain inadvertent access to sensitive data. Most commonly identified is the necessity of trust in the provider of the security measure; in this case the internet service provider which manages the malwarefilter. The majority of end-users express trust in their internet provider, but nevertheless find it difficult to evaluate the possible impact of such measures on their privacy or productivity due to a lack of transparency. Some customers express their concerns in relation to recent media attention with regards to data leaks or large-scale malpractices among technology companies that have negatively affected their customers.

> “I do not know what it does to my privacy, or where my information is sent off to. All I know is that there is a check mark and that KPN claims that makes it safer.”

In addition to requiring a degree of trust in the provider of the malwarefilter, there is a further requirement of trust in other parties affiliated with the internet service provider and the provision of the malwarefilter. Several of the interviewees mention that although they trust the ISP itself, they believe that other potentially untrustworthy parties, might also be involved in the malwarefilter’s provision. Once again, media attention on the topic of information security appears to be one of the driving factors between distrust of subcontractors or affiliations with other involved organisations.

> “I read an article about the core network the other day, where a part of the service provision is subcontracted to a different party. Which means that I can have trust in the ISP itself, but if they subcontract these things to a different party because of costs or for other reasons, and they do not do their job correctly, then you still have a problem.”

The combination of these trust issues is perhaps most prominently expressed in the perceived implications of the use of the malwarefilter for the privacy of its users. Although privacy is almost universally mentioned as an important topic in the adoption of centralised security, only some of the end-users consider there to be avoidable privacy implications associated with the use of the malwarefilter.

A significant fraction of end-users consider that although privacy is an important topic, it is not likely that there are significant privacy implications associated with the use of the malwarefilter (because of sufficient trust in the internet service provider). Others consider that privacy implications are inherently associated with the use of security measures and therefore do no consider privacy to be a significant deterring factor in the adoption of such measures.

> “It could have privacy implications if you use the data it generates for anything other than warning the customer. Let’s say, if you start commenting on the substance of users’ internet activity.”

> “Needless to say I think it could have consequences for my privacy. I think that if you want to be well protected then that inevitably comes at a cost to your privacy. I think you should simply take that for granted; that those two things are inseparable.”

Most subjects mention at least some degree of difficulty in assessing the impact of the malwarefilter or other centralised security measures on their privacy. Most customers consider there to be too little transparency about how the malwarefilter functions, or otherwise mention that it could impact their privacy depending on how exactly the measure functions.

5.1.7. Productivity and finances: the costs of a malwarefilter

The tangible costs or drawbacks associated with the malwarefilter or other security measures which are not managed by the user are mainly related to financial costs and the potential of ‘overreaction’. The expectation of financial cost is identified among approximately 20% of the subjects as a drawback that they expect to be associated with the malwarefilter, or is a cost that is otherwise often associated with the use of third-party security services. Licensing costs for the services of providers such as McAfee or Norton are mentioned as reasons to discontinue using those services, and the expectation of financial costs associated with the use of the malwarefilter is equally identifiable.

> "The malwarefilter does seem useful to me, as long as there are no extra costs attached to it."

A second, indirect form of cost that users associate with the use of centralised security services are perceived negative consequences for their productivity or the performance of their devices. Five customers note that measures which block malicious internet activity have the potential to also block legitimate traffic resulting in the loss emails or other content that merely appear to be malicious (but are not). Such measures are also to a degree expected to hamper the end-user's ability to perform other normal activities for which the devices were intended.

> "I can imagine that maybe sometimes security actually works against you. That you want to install or receive something, and first have to disable the security in order to receive it because it might appear to contain a virus for example. The negative aspects of strong security."

> "A potential drawback I see with it is that sometimes things might get blocked that do you in fact need to see. That emails that appear suspicious but are legitimate, for example emails from specific companies, get blocked while you are waiting to receive them."

5.1.8. Limited efficacy and assessing added value

Alongside the potential impact on non-malicious activities, end-users note that the malwarefilter is likely limited in its efficacy of preventing or mitigating malicious activity. These users consider, amongst other things, that attackers will simply be capable of circumventing the measure, that it is not effective against some probably threats, or that the filter used behind-the-scenes of the malwarefilter is not the best that can be obtained. Thus, the malwarefilter competes not only against the power and complexity of the threats it aims to eliminate, but also against the security measures which users themselves may employ already (four and five interviewees mention these respective aspects).

> "It is probably fine for combating the spread of malware, but I don't think it will help much against someone breaking into my account."

> "One of the disadvantages is of course that the ISP chooses the scanner that is used. It's probably fine but there are better ones out there."

Yet in the use of the malwarefilter, much like in the context of online threats and device security, one of the primary factors affecting customers' perception is their ability to assess its added value. Approximately half the interviewees mentioned that they considered it useful that their internet service provider offers security services which are managed by the ISP instead of the end-user, but that they find it difficult to imagine whether there are any benefits or potential drawbacks associated with it beyond its protective function. To nearly half the interviewed customers the functioning of the malwarefilter is difficult to comprehend and its added value difficult to assess due to a lack of information, knowledge, or visibility of results.

> It's a bit of a black box what it does. I'm not sure whether there are any advantages or drawbacks associated with using such as service; I did enable it but it provides essentially no insights into what it has done and whether and where it has brought added value.

5.1.9. Easy to manage protection: the benefits of centralised security

One of the primary benefits of the malwarefilter that was identified by six of the interviewed subjects is that it provides an added layer of security; in addition to the security measures employed by the end-user themselves. Alongside this extra layer of protection, a feeling of enhanced safety motivated five users to adopt the service.

> "It seems useful to me primarily as an addition to my own security measures; a sort of second opinion."

> "I definitely think it is a useful service because it makes me feel safer."

A third major benefit noted is fact that the malwarefilter is managed by a party that they deem to be relatively accessible (in relation to third-party products), as well as the fact that by choosing to use the malwarefilter both internet services and their protection are managed by a single, accountable party that is perceived to be (more) capable of managing them than the end-user and is potentially more accessible than providers of other security services, which are often foreign entities that do not provide extensive customer support. These potential benefits pertaining to ease of use and ease of management were most commonly cited as a benefit of centralised security; nearly a third of the respondents indicated this as a significant advantage over traditional (distributed) security solutions.

> "Until recently I had a McAfee virus application. Those are products sold by companies that you interact with once and after that you barely reach them anymore. I have the expectation that if I had to reach the internet service provider because of some issue, that that would be easier."

5.1.10. Relation to the Conceptual Model

The findings from the research interviews and subsequent thematic analysis provide some insight into the adequacy of the conceptual model in predicting the intention to use centralised DNS-based security services like the malwarefilter. The perceived (in)ability of the service to prevent malware infections and the degree to which the service is perceived to have added value in to end-users' own security configurations can all be related to the construct of perceived response efficacy. The prevalence of these concerns among customers implies that response efficacy may be a significant predictor for the intention to adopt centralised security measures like the malwarefilter. While privacy was often mentioned as a potential concern by end-users in combination with difficulty experienced in assessing the functionality of the service, many customers noted that they did not perceive there to be significant costs associated with it beyond potential financial and productivity implications.

Many of the interviewees note a lacking ability to assess the security of their internet connected devices and the possible consequences of online threats, which can be related to the constructs of perceived vulnerability and perceived severity. Information security research that employs protection motivation theory and similar models often find threat appraisal not to be a significant predictor for protective behaviour - most researchers note a relation to a lack of awareness about threats and their intangibility as a potential reason - and the interview data suggests that this might also be the case in this study. Simultaneously, the responsibility expressed by end-users for the security of their online devices seems to imply that perceived self-efficacy is high among most the study population regardless of how knowledgeable they consider themselves about cyber security.

The fact that many of the customers that had enabled the malwarefilter did so on the recommendation of the internet service provider may hint at the significance of subjective norm (beliefs about how one should behave based on the opinions and values of others) as a predictor variable. Similarly, the necessity of trust in the services' provider which was mentioned by a large fraction of subjects suggests that it may play an important role in the consideration of end-users to use centralised security services.

Section 5.2 presents the results of the questionnaire distributed among a larger set of KPN customers in order to assess the validity of the conceptual model and its explanatory power in predicting the use of centralised security measures (in this case; the malwarefilter) and the degree to which individuals engage in maladaptive coping methods through emotions such as denialism, fatalism, and wishful thinking.

5.2. Model Validation

The data used to validate the conceptual model defined in Chapter 2 was gathered using an online questionnaire distributed among roughly 1700 KPN customers with a known malware infection (malicious events that are classified as a bot or botnet, or malware) that occurred between the 1st of January 2021 and the 14th of June 2021. Additionally, the questionnaire was made available on the online forum of KPN in order to diversify the dataset. A total of 92 responses were captured and analysed. A minority of responses (14) originated from the KPN forum; the majority were obtained from customers identified through the Abuse Desk systems.

The quality of the instruments was assessed using a widely used method developed for the assessment of the internal consistency of psychometric tests; the Cronbach's Alpha metric. Cronbach's Alpha measures how closely the related items of a test are related; thus the degree to which they succeed in measuring the same 'thing'. Nevertheless, it should be taken into account that Cronbach's Alpha cannot be used to determine whether what is being measured is uni-dimensional, which requires further methods such as factor analysis. As the instrument is rooted strongly in validated theoretical frameworks and in related works, as many of its items are directly drawn from these works or adaptations of them, factor analysis is not performed.

Table 5.4 presents the Item-Total Correlations (ITC) of the survey items for each construct, and the Cronbach's Alpha (CA) value of the composite constructs of perceived severity (PS), perceived vulnerability (PV), perceived response efficacy (PRE), perceived response costs (PRC), perceived self-efficacy (PSE), and subjective norm (SN). Table 5.5 displays the correlation matrix of the composite constructs. Correlation values close to 1 may indicate that constructs implicitly measure the same concept; values closer to 0 are more desirable given the assumed independence of model constructs.

Table 5.4: Instrument quality assessed through item-total correlation and Cronbach's Alpha.

Item	ITC	CA	Item	ITC	CA	Item	ITC	CA	Item	ITC	CA
PS1	0.562	0.813	PRE1	0.828	0.890	PSE1	0.606	0.823	MC1	0.381	0.496
PS2	0.743		PRE2	0.769		PSE2	0.718		MC2	0.333	
PS3	0.597		PRE3	0.760		PSE3	0.724		MC3	0.250	
PV1	0.369	0.629	PRC1	0.731	0.777	SN1	0.535	0.661	TR1	0.635	0.768
PV2	0.474		PRC2	0.633		SN2	0.565		TR2	0.635	
PV3	0.479		PRC3	0.493		SN3	0.348				

Table 5.5: Correlation matrix of predictor variables.

	PS	PV	PRE	PRC	PSE	SN	TR
PS	1.000						
PV	-.067	1.000					
PRE	-.027	0.131	1.000				
PRC	0.110	-.002	0.177	1.000			
PSE	-.119	-.032	-.048	0.270	1.000		
SN	-.193	0.009	0.216	0.113	-.317	1.000	
TR	-.123	-.210	-.318	0.044	0.003	-.171	1.000

5.2.1. Subject Demographics

Respondents were predominantly middle-aged and male (almost three quarters of all respondents); three respondents reported either a not conforming to being either male or female, or decided not to provide this information. Higher education represents approximately half the respondents (HBO or WO in the Dutch education system or equivalent), with a further 33% of respondents indicating they have received a vocational education (MBO in the Dutch education system or equivalent), with a small minority of approximately 12% of respondents having received at most a secondary education (Voortgezet onderwijs, mavo, havo or vwo in the Dutch education system, or equivalent). Figure 5.1 presents a visualization of the age distribution and education levels among male and female respondents in both the sample and the KPN customer base. Figure 5.2 presents a visualization of the distribution of highest achieved level of education between male and female respondents, and concern about online threats between IT-professionals and non-IT-professionals.

Nearly 40% of respondents had either received an education in the field of information technology, are actively employed in the IT sector or have previously been employed there. 15% of end-users consider themselves inexperienced in the use of technology, while roughly 60% consider themselves experienced, and a further 25% consider themselves highly experienced users of technology. Thus, a significant fraction of respondents are users that - either through their education or profession, or through private activities - consider themselves to have a high level of affinity with technology. Despite this apparent familiarity with information technology relatively few respondents indicated a significant level of concern about online threats, although those with an IT-related background expressed with relatively greater frequency that they were either not concerned or very concerned.

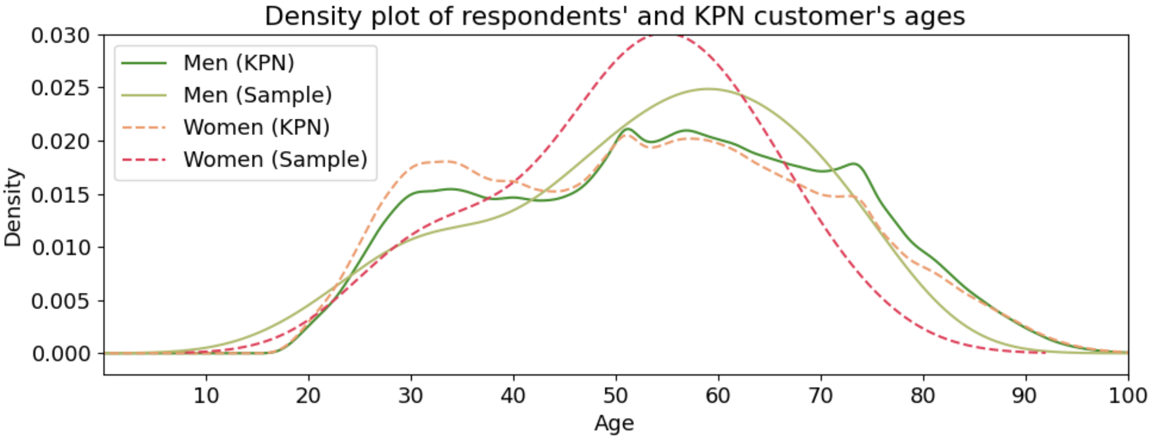


Figure 5.1: Age distribution among survey respondents and the KPN customer base.

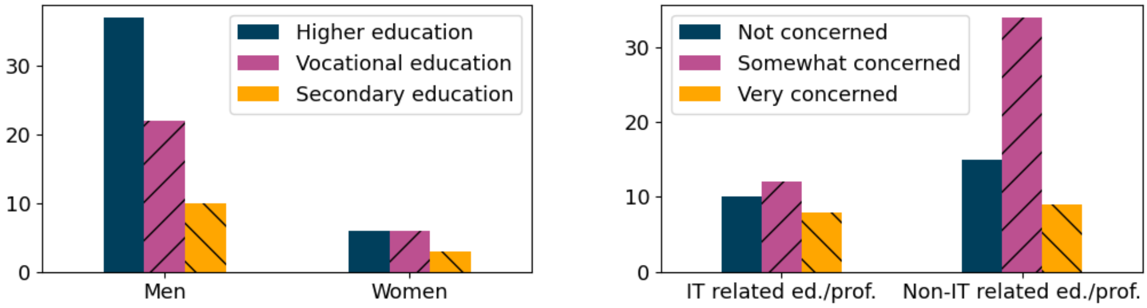


Figure 5.2: Distributions of education relative to gender, and concern about online threats with regards respondents having enjoyed an IT-related education or being employed in an IT-related profession.

Device Ownership & Use

Respondents were inquired about ownership and usage of three types of devices: (1) traditional (non-mobile) computing devices such as desktop computers and game consoles, (2) mobile devices such as laptops, tables, and smartphones, and (3) smart-home devices such as smart speakers, intelligent thermostats, and smart lighting. With the exception of a single case, all respondents noted ownership of both traditional computing devices and mobile devices. Ownership of traditional computation devices was similarly widespread albeit slightly less than mobile devices. Approximately half the respondents reported owning at least one smart-home device.

Figure 5.3 provides an overview of the self-reported frequency at which respondents interact with these various types of devices. Traditional and mobile computing devices are overwhelmingly used on a daily basis, although roughly 15% of respondents indicate that they use these devices only on a weekly basis or even less frequently in the case of traditional computers, while mobile devices are used on a daily basis virtually without exception among the respondents. Approximately half of the respondents indicated that they own at least one smart-home device. Similar to other computing devices, smart-home devices are used on a daily basis by most users although a notable fraction of users (about one-third of smart-home device owners) reported using their devices on a weekly or monthly basis or less frequently.

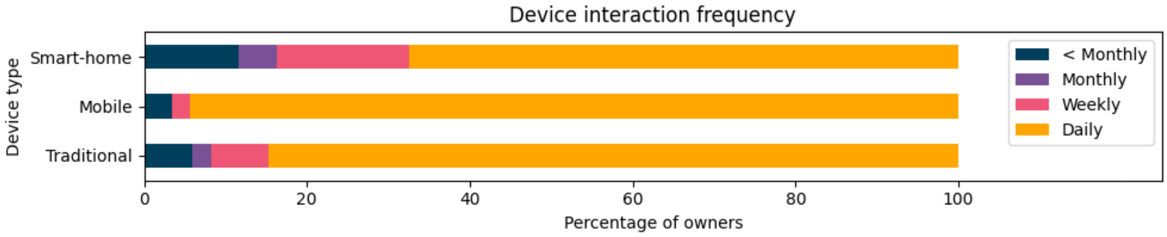


Figure 5.3: Device interaction frequency.

Security incidents & Practices

When asked about the extent to which respondents are concerned about the dangers posed by online threats, approximately 75% of the respondents reported being concerned about the dangers posed by online threats with roughly a third of them indicating that they were very concerned. The remaining 25% reported that they were not at all concerned about online threats.

With regards to previous experiences with security incidents, 18% of respondents reported having had no previous experience with security incidents. Of the remaining subjects almost everyone reported having previously run in to phishing or other forms of deceptive messages. Previous experience with malware incidents were reported by half the respondents, with data theft or other forms of known, significant data compromise being a seemingly rare occurrence among less than 10% of the subjects.

Figure 5.4 gives some insight into the frequency with which respondents perform certain security actions. Security updates are a common practice among the queried subjects, which 40% report doing often and another 40% report doing regularly. Roughly 10% of respondents report performing security updates either seldom or never. Changing passwords is significantly less common as a security practice, with only 25% of end-users reporting that they perform such actions either regularly or often. A third of respondents change their passwords occasionally, with the remainder (almost 40%) reporting that they change their passwords either seldom or never.

Of the types of security software and measures used, third-party security services are most commonly reported (over half of the respondents use third-party software). A little under half the respondents use built-in or default security measures of their devices or the operating systems they run such as Microsoft Defender. Approximately 15% of respondents report the use of security measures provided by their internet provider itself, such as KPN veilig or the KPN malwarefilter, with roughly 10% reporting that they do not employ any of these security measures.

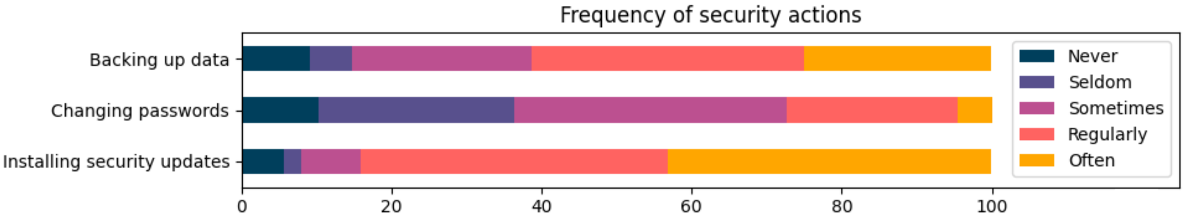


Figure 5.4: Security action frequency.

Malwarefilter

Lastly, customers were inquired about their familiarity with - and use of - the KPN malwarefilter and, after being presented with a short explanation of its functioning in case they are not, queried whether they would consider using the malwarefilter in the future. Approximately 64% of the respondents indicated that they were not familiar with the malwarefilter, while 36% were familiar with it. Of those who were familiar with the service a quarter indicated that they have the service enabled, while half of them indicated they do not use the service, and another quarter indicated that they were not sure whether they had it enabled or not.

All respondents were asked how likely they were to either enable or keep using the malwarefilter in the future. More than half indicated that they were likely to enable or keep using the service, alongside 10% of respondents who indicated that they were very likely to enable or keep using it. Approximately 20% of respondents indicated that they were unlikely to use the service, alongside 10% who indicated that they were very unlikely to do so.

5.2.2. Intention to Use the Malwarefilter

Binary logistic regression was performed to evaluate the adequacy of the conceptual model in predicting attitude towards the use of centralised security measures. While intention to use the malwarefilter was measured on an ordinal scale, it was deemed more appropriate to convert this scale to binary in light of the limited number of responses that could be collected. Respondents that indicated being either unlikely or highly unlikely to enable the service are considered to have no intention to adopt the service, while those who indicated being either likely or highly likely to enable the service were deemed to have an intention to enable the service. The proposed model is significant at $X^2(7), p < 0.01$, explaining approximately 31% to 44% of the variance in the intention to use the malwarefilter.

Significance of the constructs is determined using three common threshold values: $p < 0.10$, $p < 0.05$ and $p < 0.01$. Table 5.6 denotes the significance of each construct in predicting the intention to use the malwarefilter. Neither perceived severity nor perceived vulnerability were found to be significant predictors for the intention to adopt centralised security measures. Perceived response costs, alongside trust in the internet service provider, were similarly found to not be significant predictors despite the fact that financial and productivity impact of enabling the malwarefilter were cited by a number of interviewees as perceived drawbacks of the service. Perceived response efficacy was found to be significant at $p < 0.01$ alongside self-efficacy and subjective norm at $p < 0.05$.

The positive B-value associated with perceived response efficacy and indicates that as subjects consider the malwarefilter to be more effective at mitigating malicious activity, they are more significantly more likely to display an intention to adopt the measure. At an Exp(B) value of 5.880 the effect can be considered to be large. Similarly, the positive B-value associated with subjective norm and the Exp(B) value of 3.966 indicates that as subjects consider online safety to be an important topic in their social or professional life or to authoritative parties such as their internet provider, they are significantly more likely to have an intention to adopt the malwarefilter. The Negative B-value associated with self-efficacy indicates that end-users who consider themselves more capable of successfully employing security measures are significantly less likely to have an intention to use the malwarefilter.

Relation to Literature

These findings are largely in line with those of earlier works examining the adequacy of the PMT model and its derivatives in examining the adoption of information security practices and tools. Perceived severity and perceived vulnerability are at best found to be inconsistent predictors Mills & Sahi (2019), often as a result of lacking awareness or knowledge of information security and the general complexity and intangibility associated with cyber security and cyber threats (Liang & Xue, 2010). Yet, response efficacy is nearly universally found to be a significant predictor among studies that employ the construct, indicating that there is no apparent link between the use of security measures and whether end-users perceive threats to be both a realistic and impactful to themselves or others; as long as they think a countermeasure is generally effective they might be willing to enable simply to prevent as much risk as possible. The findings of the thematic analysis underwrite the importance of perceived response efficacy in the intention to use the centralised security measures, as added value and limitations in the protective ability of the service are often cited as reasons not to enable the service or, conversely, as reasons to enable it in case the end-user perceives that there is added value in enabling it.

Response cost is occasionally found to be a significant predictor for intention to adopt security measures among studies involving home-users, but this is not the case in this study and in some previous works on home-user security behaviour (Menard et al., 2017; Mills & Sahi, 2019). Earlier works have largely considered perceived costs to be primarily related to the financial implications of the use of security tools and behaviours, which are not present in the case of the malwarefilter (as it is a free service) and might therefore explain the lack of predictive power of this construct. This is affirmed by Hanus & Wu (2016) who state that the degree to which response costs is found to be significant might be strongly impacted by respondents' expectations of services being provided for free (e.g. free or freemium software, or bundled with the sale of other products). Findings from the thematic analysis are in slight support of these quantitative results, as - while some interviewees mention that financial costs or productivity impact could be a deterrent - most customers do not associate notable disadvantages with the use of centralised security services. This perception may be (partially) explained by the degree of trust expressed in the ISP by most users.

5.2.3. Maladaptive Coping with Emotions

Linear regression was performed to evaluate the adequacy of the conceptual model in predicting maladaptive coping using three classes of emotions (denialism, fatalism, and wishful thinking) in line with the recommendations by Haag et al. (2021). Denialism pertains to the conscious or subconscious denial of the reality of threats; preferring not to think about them rather than to face them. Fatalism refers to the idea that individuals might feel that they are subject to faith and that their actions do not affect the outcome of events. Wishful thinking is the idea that an individual prefers an (unattainable) situation where the threats do not exist. Few prior efforts to study information security behaviour using PMT-based hypotheses have examined maladaptive coping (only three out of 61 studies identified by Haag et al.), none have examined maladaptive coping with emotions. The proposed model is significant at $X^2(7), p < 0.01$, explaining approximately 27% of the variance in maladaptive coping with emotions.

Table 5.6: Significance and effect size of model constructs in predicting intention to use the malwarefilter and maladaptive coping with emotions. Significance levels are denoted as *, **, and *** for p-values <0.10, <0.05, and <0.01 respectively. B-values denote the direction of the effect, Exp(B) and Standardized B are indications of the effect size.

Intention to Use			
Construct	p-value	B	Exp(B)
PS	0.706	-0.174	0.840
PV	0.576	0.213	1.237
PRE	0.001***	1.772	5.880
PRC	0.853	0.075	1.078
PSE	0.025**	-1.077	0.340
SN	0.026**	1.378	3.966

Maladaptive Coping			
Construct	p-value	B	Standardized B
PS	0.398	-0.088	-0.085
PV	0.015**	0.242	0.245
PRE	0.046**	0.172	0.201
PRC	0.000***	0.346	0.396
PSE	0.560	-0.058	-0.062
SN	0.489	0.088	0.075

Significance of the constructs is determined using three common threshold values: $p < 0.10$, $p < 0.05$ and $p < 0.01$. Table 5.6 denotes the significance of each construct in predicting maladaptive coping with emotions. Similar to the intention to use the malwarefilter, perceived severity was not a significant predictor of maladaptive coping with emotions. However, perceived vulnerability is significant at $p < 0.05$ alongside perceived response efficacy (also at $p < 0.05$), and perceived response costs at $p < 0.01$. Neither trust or self-efficacy, nor subjective norm were found to be significant predictors for maladaptive coping with emotions.

These findings suggest that end-users tend to engage in maladaptive coping with emotions significantly more frequently as they perceived themselves to be more vulnerable to online threats, as they perceive the response to be more effective at mitigating threats, and as they perceive the costs of implementing countermeasures to be higher. Of these constructs, the perceived costs of employing the security measure is the strongest predictor for engaging in maladaptive coping with emotions, followed by perceived response efficacy and perceived vulnerability.

Relation to Literature

As users consider themselves more vulnerable to such threats, they may attempt to cope by denying the reality of the threats or convincing themselves that ‘there is nothing that could have been done’, as indicated by the significance of perceived vulnerability. Surprisingly, perceived response efficacy is found to be a significant predictor for maladaptive coping with emotions. A possible explanation of this effect may be found in the existence of denialistic or fatalistic tendencies on the side of the end-user, that may be reduced through the intervention of an authoritative party. In other words, while online threats may be perceived to pose a danger that the individual has only limited ability to affect or attempt to avoid, supposedly more knowledgeable parties such as an ISP may nevertheless be able to provide effective protection beyond the capabilities of the end-user. The absence of self-efficacy as a predictor for maladaptive coping behaviour suggests that end-users do indeed experience a limitation in their ability to protect their devices; a limitation which they may perceive not to exist for other parties.

The significance of perceived response costs in predicting maladaptive coping with emotions indicates that individual who think their security comes at a great cost - whether financial, or in time or effort require to enable or use - are more likely to deny the reality of threats, consider them unavoidable, or engage in wishful thinking about realities where these security measures would not be required. The existence of this relationship in the context of its absence in the intention to adopt the malwarefilter suggest that if individuals perceive its costs to be higher they might resort to justifying its use or disuse through maladaptive coping with emotions rather than choosing to use or not to use the measure as a result. As noted by (Haag et al., 2021), maladaptive coping does not necessarily suppress or counteract adaptive coping behaviours.

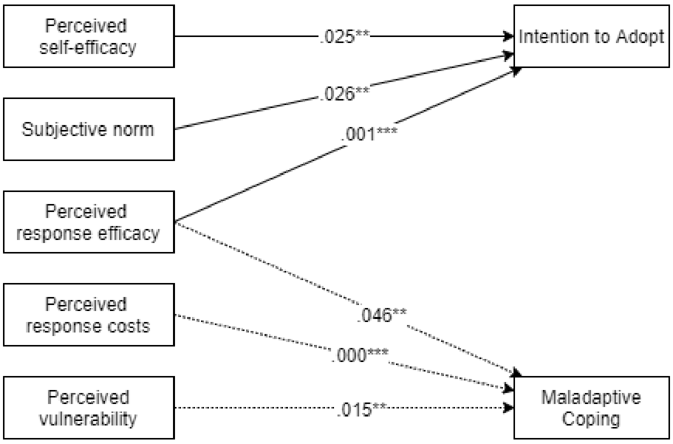


Figure 5.5: Significance of the model constructs in predicting intention to adopt the malwarefilter and maladaptive coping with emotions.

5.3. Malwarefilter Efficacy

Malware infections were monitored for 284 customers who were contacted about a malware infection and the malwarefilter (and to whom this notification was successfully delivered) in the months of February and March 2021. These customers received an email at the start of April notifying them of the malicious activity and inviting them to enable the malwarefilter. Two monitoring periods can be distinguished; the period before the malwarefilter notification was sent, constituting the months of January, February, and March 2021, and the period after the malwarefilter notification was sent; the months of April, May, and June of that same year.

Use of the malwarefilter among these 284 users was assessed by querying their IP addresses in a platform that records the DNS queries filter-users, these IPs were queried one week after the notification had been sent. Among the 284 contacted individuals, 25 individuals were identified as having enabled the malwarefilter at the point assessment defined earlier, an approximate 8.5% of the contacted users (roughly 40 times more than among the general population at the start of the experiment period, according to estimates by KPN).

5.3.1. Malware Activity

Figure 5.6 present an overview of the dates at which the identified malware infections are expected to have occurred; the date where evidence of the infection was first found or reported. A total of 776 malware infections were identified from January 1st to June 30th among the individuals enrolled in the experiment. 525 of these infections occurred in the first monitoring period (before the malwarefilter notification), 251 infections occurred in the second monitoring period (after the notification).

In addition to the notably lower number of infections in the second period relative to the first period, the first period is characterized by a significant spike in overall malware activity in early February while such peaks are absent during the second monitoring period. Overall, malware activity appears to be more constant during the second period, while there is a notable build-up and cool-down before and after the activity spike in the first period (with especially little activity recorded in March).

Despite these notable differences in the number of infections occurred within a period, there appears to be less difference in the duration of infections between the periods (Figure 5.7). During both monitoring periods, the majority of infections are resolved (that is; evidence for their existence is no longer found or reported) within a week, with a large fraction of resolutions occurring within the first day of infection. Nearly all infections are resolved in a little over two weeks from the date of infection, while a handful of infections are only resolved after a month or longer.

5.3.2. Infection Remediation

Infection survival rate was analysed using Kaplan-Meier survival analysis. As the absence of secular trends cannot be assumed (one of six assumptions underpinning Kaplan-Meier survival analysis), the survival functions of the first and second monitoring period are not directly comparable. Nevertheless, they provide a visual aid in understanding the survival rate of malware infections over time and the differences between both monitoring periods. Figure 5.8 displays the survival functions of the three groups of subjects (not notified about the malwarefilter, notified but did not enable the malwarefilter, notified and enabled the malwarefilter) during the first and second monitoring periods.

Similar remediation patterns can be observed among customers that received the email notification but did not enable the malwarefilter across both monitoring periods, although infection remediation times are notably longer for the majority of infections in the second monitoring period as can be deduced from the overall higher survival

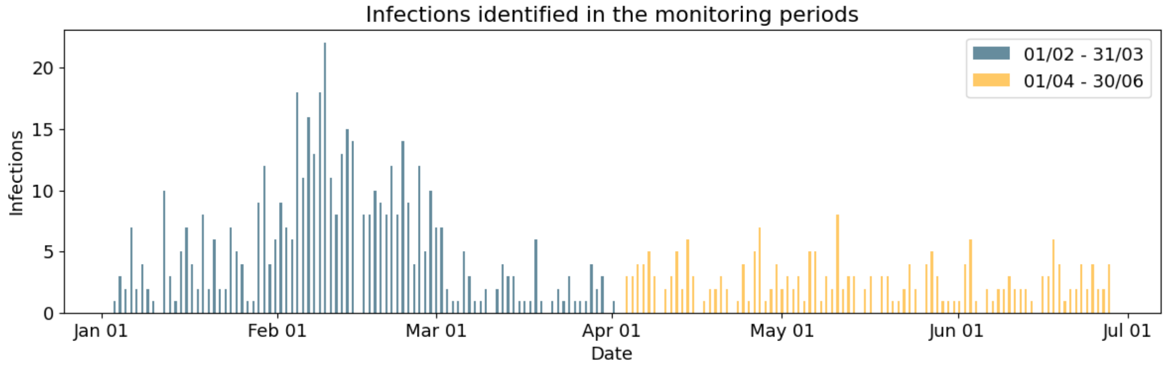


Figure 5.6: Date at which an infection was first identified for infections between January 1st and June 30th of 2021.

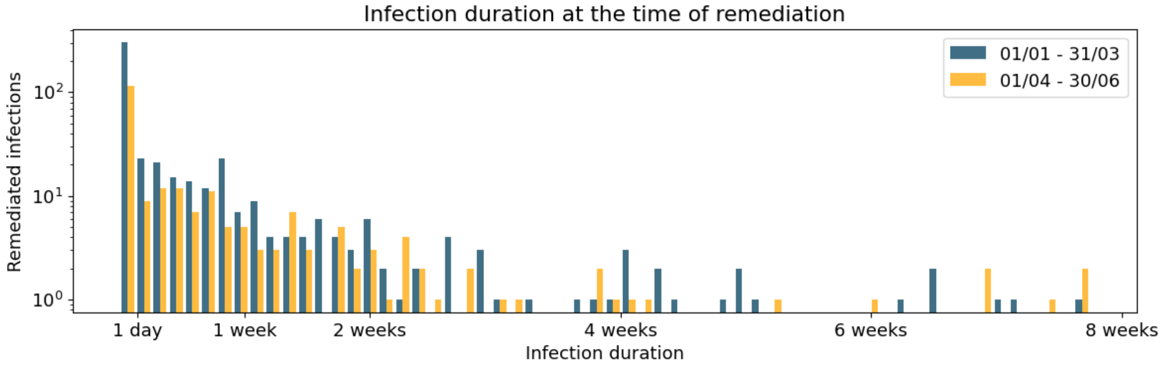


Figure 5.7: Infection duration in days for infections that occurred between January 1st and June 30th of 2021.

levels particularly between 100 and 800 hours since first evidence. Infection survival patterns for customers that did not receive the communications about the infection and the malwarefilter also show notably greater survival at the interval between 0 and 200 hours since first evidence. A significant difference can be observed in the survival patterns for those that did enable the malwarefilter in response to the notification, although this may be attributed to the low number of infections among this group in the second monitoring period.

5.3.3. Malware Types

Figure 5.9 provides an overview of the frequency at which various types of malware are encountered among the infection cases, and the differences in these frequencies between the two monitoring periods. Among both the first and second monitoring period, a number of malware types occur with great frequency. Notably, three different strains of the Avalanche malware family make up a large fraction of the identified infections. Alongside the strains of Avalanche, the Bladabindi, Gamarue, Necurs, Sality, and Qrypter.rat malwares occur fairly often among the infection cases in both the first and second monitoring period. Other types of malware such as Conficker/Downadup, Bankpatch and Rovnix appear only or predominantly during the first or second period but not in the other. The (lack of) presence of these specific families during either period is likely the result of the limited sample size, as their occurrence mostly pertains to one or two infections during the given monitoring period(s).

Avalanche is a botnet that had in 2016 been dismantled by an international operation against the criminal syndicate that operates it. It infects computers running Microsoft Windows, employing compromised hosts for a variety of tasks such as email spam and phishing attacks, as command and control servers for other malwares, and more. While the total number of occurrences of the avalanche malware are approximately equal across both periods, a notable difference can be identified between the prevalence of the generic strains and the Andromeda strain, which is much more prevalent in the second monitoring period.

Gamarue and Bladabindi, similarly, primarily target the Windows operating system in an effort to allow sustained remote access to compromised devices and their data. The Bladabindi malware is significantly less prevalent during the second monitoring period, while the relative frequency of Gamarue infections has increased approximately proportionally to this reduction. Conficker/Downadup also target Windows-operated systems in an effort to install secondary payloads such as spambots and scareware. Qrypter is malware-as-a-service, remote

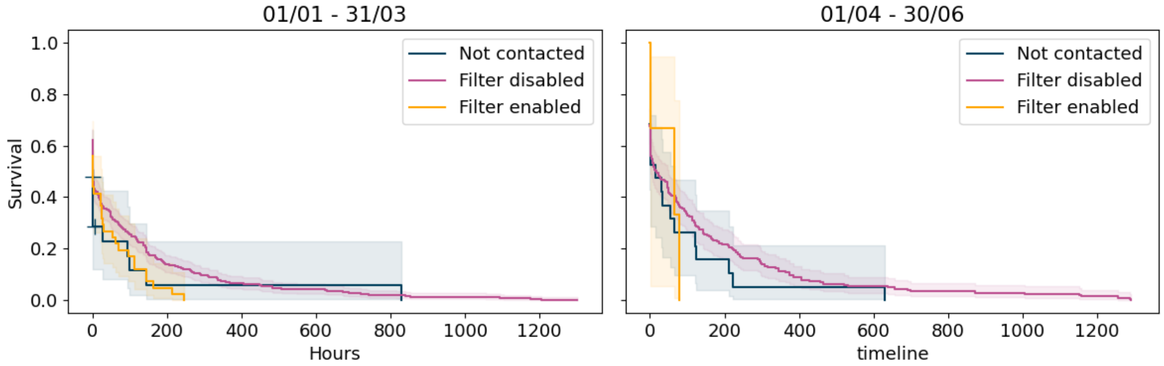


Figure 5.8: Kaplan-Meier survival analysis.

access Trojan leveraging TOR-based command and control structures.

Qsnatch and Mirai, unlike almost all of the other malwares that were detected, specifically target systems other than (Windows-operated) desktop computers. Both Qsnatch and Mirai are well-known IoT malware families and Mirai specifically has previously been used in some of the most notorious DDoS attacks. While mirai maintains a presence of approximately 5% of the identified infections during both monitoring periods, Qsnatch infections were only identified among the monitored subjects during the second monitoring period. The overall prevalence of IoT malware among the identified infections is relatively low considering that the questionnaire demographics indicate approximately 50% ownership of IoT devices, which may be an indication that conventional computation devices such as desktop and laptop computers still pose a significantly greater threat to cyber security than these new devices.

5.3.4. Malwarefilter Impact

In order to gauge the significance of the observed differences and the degree to which they may be accounted to the malwarefilter or simply be the result of changing circumstances or other interventions, a number of tests are conducted for statistical significance. These hypothesis tests are presented in Table 5.7 and conducted between the two groups that had received the malwarefilter notification so as to account for the effect of the notification itself. Based on the tested hypotheses, a number of statement can made about the perceived impact of the malwarefilter on malicious activity among the subjects:

Table 5.7: Tested hypotheses about the number of infection cases and their duration.

	Null Hypothesis (H_0)	**Status**	**p**
H1a	There is no significant difference in the distribution of malware infections across users and non-users of the malwarefilter before the intervention.	Retain H_0	0.812
H1b	There is no significant difference in the distribution of malware infections across users and non-users of the malwarefilter after the intervention.	Reject H_0	0.039
H2a	There is no significant difference in the distribution of infection durations across users and non-users of the malwarefilter before the intervention.	Retain H_0	0.475
H2b	There is no significant difference in the distribution of infection durations across users and non-users of the malwarefilter after the intervention.	Retain H_0	0.116
H3a	There is no significant difference in the distribution of malware infections among users of the malwarefilter before and after the intervention.	Reject H_0	0.002
H3b	There is no significant difference in the distribution of infection durations among users of the malwarefilter before and after the intervention.	Reject H_0	0.000
H4a	There is no significant difference in the distribution of malware infections among non-users of the malwarefilter before and after the intervention.	Reject H_0	0.000
H4b	There is no significant difference in the distribution of infection durations among non-users of the malwarefilter before and after the intervention.	Reject H_0	0.000

- Significant differences in both the occurrence of infections and their duration were observed among all subjects that had received the notification regardless of whether they enabled the malwarefilter or not. A notably

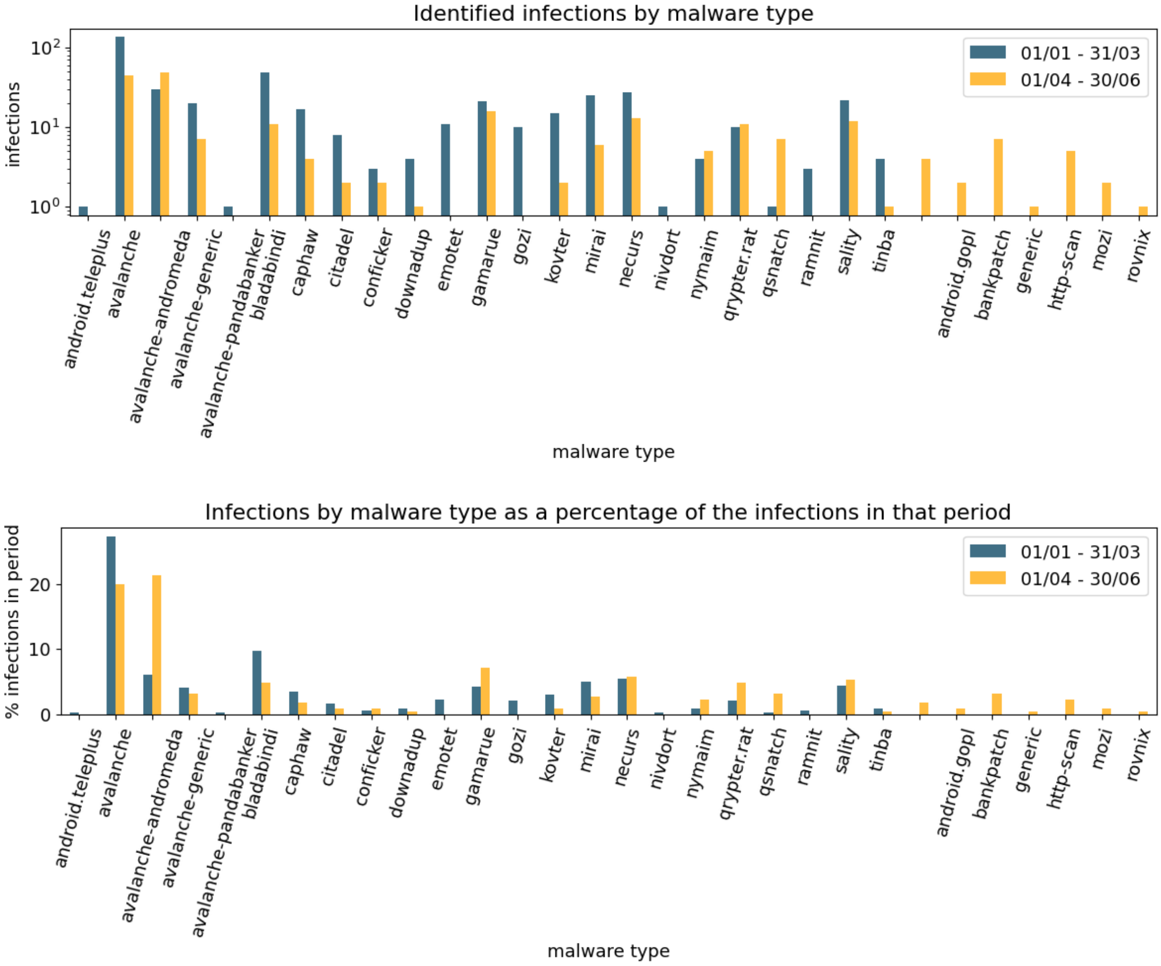


Figure 5.9: Malware types identified among infections between January 1st and June 10th of 2021 by absolute number (top) and percentage of infection within the associated monitoring period (bottom).

smaller reduction in vulnerability is observed among those customers that did not receive the notification at all, indicating the notification itself elicits some protective behaviour or awareness.

- There was no significant difference between the infection characteristics of the customers that had enabled the malwarefilter and those that had not before the notification was sent, neither in the number of infections that occurred nor in their duration, indicating a level playing ground at the start of the experiment.
- A statistically significant difference was observed in the second monitoring period between customers that had enabled the malwarefilter and those that had not in terms of the number of infections but not in their duration, indicating that the intervention significantly reduces end-users' vulnerability to infections but does not significantly affect remediation time.

Table 5.8 provides an summary overview of the number of infection cases, their average duration, and the percentual change in infections between the first and second monitoring periods. In the period before the introduction of the intervention (from January to March of 2021), 525 had been identified across the three subject groups. 55 infections had been identified among customers that would not receive the malwarefilter notification in April, 429 among customers that would receive the notification but not enable the malwarefilter, and 41 among the customers that would receive the notification and enable the malwarefilter.

In the second monitoring period (April to June of 2021) a total of 251 malware infections were identified across all subject groups (less than half the total number of infections relative to the first period). Of these 251 infections, 45 were identified among the 29 customers that did not receive the malwarefilter notification. 203 infections were identified among the 259 customers that received the notification but had not enabled the malwarefilter, while 3 infections were identified among the 25 customers that had enabled the malwarefilter.

Simultaneously, among both the users that enabled the malwarefilter and those that didn't an increase in the remediation time for malware infections can be observed. The average duration of a malware infection in the first monitoring period was approximately 87 hours, while this average duration increased to 127 hours in the second monitoring period. The total number of identified infections among the subject groups was reduced by approximately half in the second monitoring period, likely as a combined result of existing remediation processes, the interventions of this experiment, and shifting activity patterns.

Table 5.8: Infection occurrence and average duration in hours during the first and second monitoring periods.

Group		Before Intervention		After Intervention		Percentual Change	
Malwarefilter	**Size**	**Infections**	**Avg. Dur.**	**Infections**	**Avg. Dur.**	**Infections**	**Avg. Dur.**
All	313	525	87.64	251	126.72	-56.17%	+35.72%
Not notified	29	55	63.47	45	88.05	-18.18%	+38.73%
Disabled	259	429	94.24	203	132.28	-52.68%	+33.58%
Enabled	25	41	38.40	3	48.92	-92.68%	+27.40%

5.3.5. Relation to the Research Interviews and Conceptual Model

The findings presented in the previous paragraphs can be related to the research interviews and conceptual model in a few noteworthy ways. Firstly, the strong reduction observed in the number of infection cases among both subject groups that received the malware and malwarefilter notification relative to the customers that did not receive the notification indicates that fear appeals alone are a potentially strong motivator to engage in one form of adaptive coping or another. The large reduction in infection cases among those who enabled the malwarefilter seems to therefore be in part related to heightened security awareness or (temporary) improvements in security behaviour. Secondly, most of the interviewed customers that had enabled the malwarefilter owned at least one smart-home device, and the survey responses similarly indicate a reasonably high degree of ownership of IoT devices among KPN customers.

Despite this, malware families that target IoT devices were relatively rare among the monitored infections, indicating that traditional computation devices still form a major cyber security threat and infection opportunity for botnets, perhaps more so than IoT devices in the case of most end-users. Thirdly, the significance of self-efficacy as a negative predictor for intention to use centralised DNS-based security in combination with the high rate of reinfection among customers that did not enable the malwarefilter provides evidence that many users who consider themselves capable of protecting their online devices may be severely overestimating their ability to do so. Lastly, the fact that remediation time is seeming unaffected by the use of the malwarefilter is line with statements made by interviewees about the lack of feedback to the user and a lack of transparency about its functioning as a deficiency of the service.

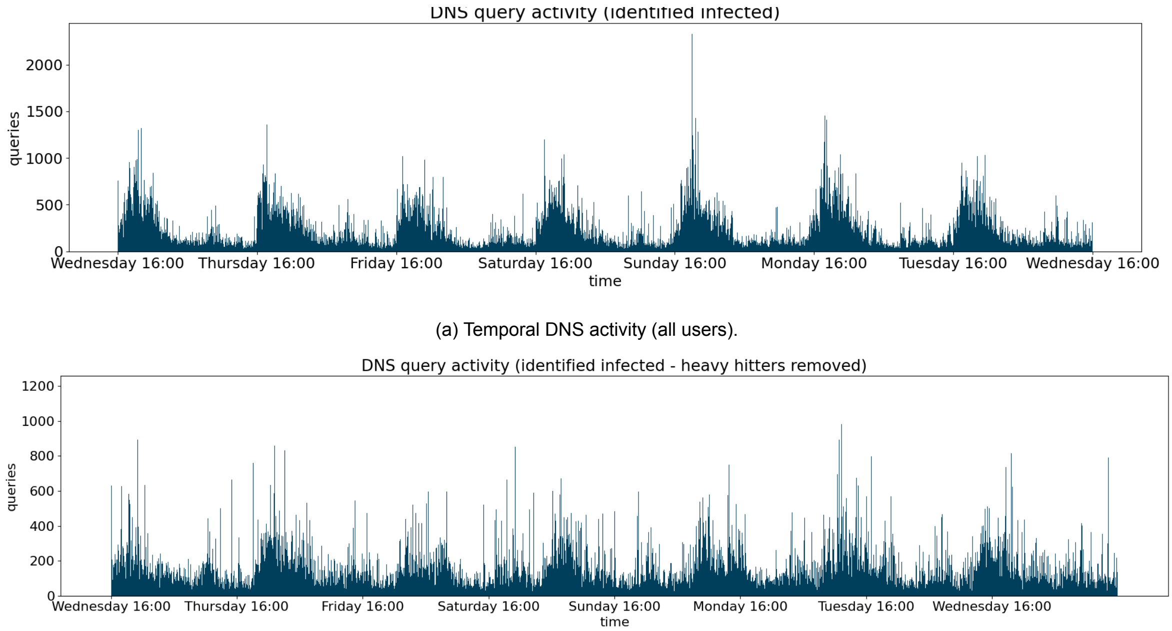
5.4. DNS Exploration

Operating the malwarefilter offers opportunities to enhance mitigative capability through the analysis of DNS traffic captured by the service. Earlier works have shown that a variety of statistical and machine-learning based techniques have the potential to detect malicious DNS queries based on traffic characteristics. Successful exploitation of this data may thereby offer protection beyond simple blocklisting methods, notably by allowing for the detection of compromised devices without the need for external reporters beyond the construction of a comparison dataset. Previous works most commonly employ temporal activity and query features such as domain name composition to distinguish legitimate and illegitimate traffic.

To retrieve the DNS data, a PowerDNS-based platform is queried for requests made by a specific set of IP addresses during a set period of time. As this platform is not set up to perform systematic monitoring but rather as a troubleshooting aid there are limitations associated with the data retrieval process. Notably, data has to be retrieved manually and for each IP address individually. In order to provide an as-complete-as-possible overview of the temporal DNS activity while accounting for the limitations imposed by the manual nature of the process, the DNS platform is queried for the last 24-hours of DNS activity of each IP address on 16:00. Up to 30,000 records can be achieved for each individual IP address and each 24-hour period, which nevertheless imposes a limitation on the completeness of the data for users who make (much more) than 30,000 DNS requests per day. Downtime of the platform and other miscellaneous issues associated with the nature of the platform impose further restrictions on the data that could be obtained. As such, the explorations primarily concern data from the period of May 19th to May 26th of 2021; the longest consecutive period for which data is considered to be complete.

5.4.1. Temporal Activity Patterns

Figure 5.10a presents an overview of the accumulated query activity of the monitored users throughout the aforementioned period. There are major activity spikes in the number of DNS queries made from roughly 16:00 to 20:00 where activity starts to diminish until the early morning. Throughout most of the day a relatively stable activity pattern can be observed with local spikes typically around mid-day, until the activity cycle starts anew around 16:00. Activity remains fairly consistent throughout the week, displaying no notable activity increases or decreases during specific days or during the weekend. Figure 5.10b presents the same dataset with the exception of a number of ‘heavy hitters’; end-users whose DNS activity significantly exceeds the limit of 30,000 retrievable records per day and whose inclusion therefore skews the pattern towards the beginning period of each day (where a day is the 24-hour period from 16:00 to 16:00, as delineated earlier). Query activity patterns for individual users can be found in Appendix H.



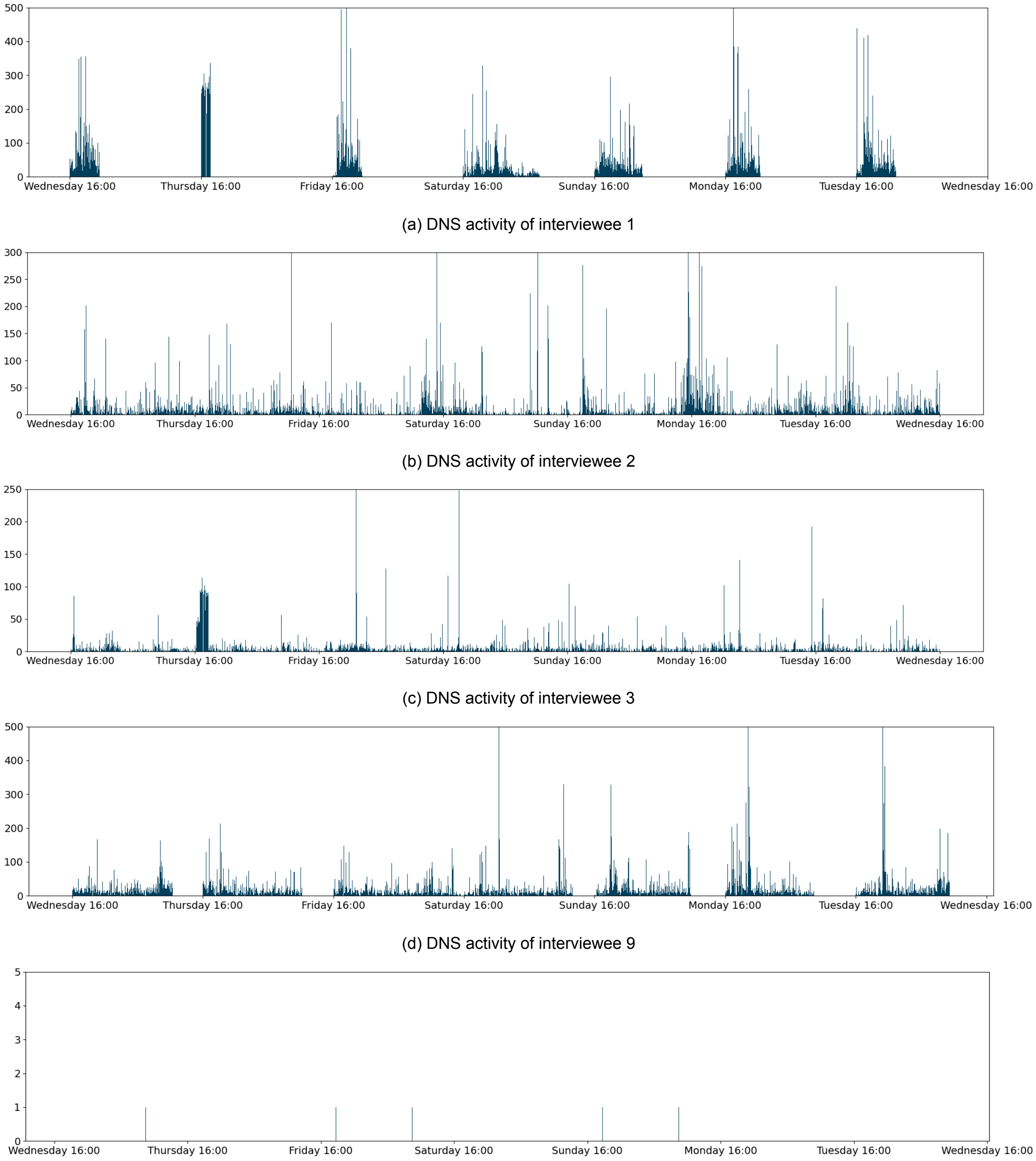
(a) Temporal DNS activity (all users).

(b) Temporal DNS activity (heavy hitters removed).

Figure 5.10: Temporal DNS activity of monitored users from May 19th to May 26th.

Interviewed Customers

Of the 25 customers that enabled the malwarefilter nine had participated in the research interviews; interviewees 1, 2, 3, 6, 9, 10, 15, 16 and 20. DNS log data could not be retrieved for interviewees 9 and 20; the activity patterns of the other interviewees correspond to those presented in Figure 5.11. All three identified malware infections among malwarefilter subscribers occurred among subjects that were not interviewed.



(a) DNS activity of interviewee 1

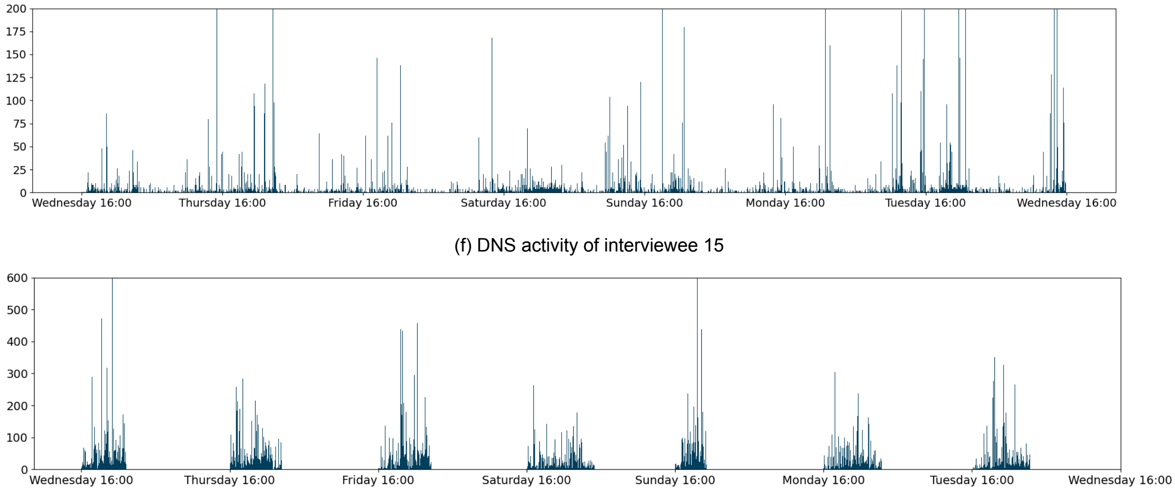
(b) DNS activity of interviewee 2

(c) DNS activity of interviewee 3

(d) DNS activity of interviewee 9

(e) DNS activity of interviewee 10

DNS queries originate from the IPs of interviewees 1, 10 and 16 only during specific times of day, whereas the activity patterns of other interviewees display significant background activity. Users 1, 2, 3, 9 and 15 indicated owning one or more smart-home devices, while users 10 and 16 did not. IoT devices are typically always-on (which is part of the reason that these devices are attractive targets for threat actors) and may therefore be expected to produce a degree of activity during hours that most users generally would not be actively using their devices. Alternatively, smartphones are kept in an always-on state by most of their owners and may explain some of the traffic generated during non-active hours.



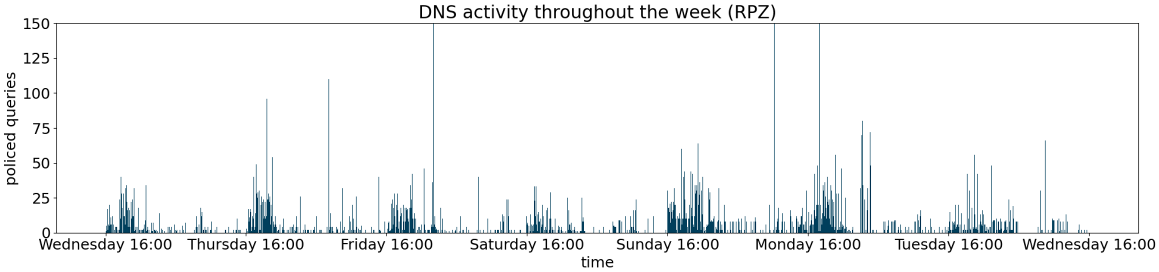
(f) DNS activity of interviewee 15

(g) DNS activity of interviewee 16

Figure 5.11: Temporal DNS activity of interviewed end-users.

Furthermore, the activity graph of interviewee 10 (Figure 5.11e) displays only a handful of queries made throughout the monitoring period, which is indicative that the vast majority of DNS requests made by this user's devices are not actually routed through the DNS servers that host the malwarefilter. This interviewee noted that among the software and security measures they employ is a Virtual Private Network (VPN). VPN services - especially the most common premium ones - often allow their users to not only tunnel their regular internet traffic through the VPN but also their DNS traffic. The occasional queries might in turn be explained by a device that makes periodic requests, for example to synchronise time or check for updates, which does not operate across the VPN.

Approximately 2% of the queries in the monitored period were blocked in accordance with a response policy zone, indicating that the answer returned to the client was modified. This may indicate blocking a page request because of its presence on a malware or phishing blacklist, but may also indicate modifications for other reasons. Figure 5.12a provides an overview of the temporal activity of DNS queries that were tagged in accordance with these policy reasons. Several significant peaks in the number of queries caught by the RPZ can be seen at various points throughout the week, notably at two short time windows on Friday evening and at several points of time on Monday.



(a) DNS queries tagged with a policy reason throughout a single week.

Known-infected Customers

Due to the limited number of infections among the experiment subjects that enabled the malwarefilter, a second set of DNS records were obtained for 13 IP addresses known to have made one or more requests for malware-associated domains within the period of Thursday June 3th to Wednesday June 9th 2021. This period is one day shorted than the full week overview presented in the previous figures and excludes data from the 24-hour window between Saturday 16:00 and Sunday 16:00 due to problems retrieving log data for the monitored IP addresses on these days. Figure 5.13 visualizes DNS query activity for these 13 customers during the aforementioned period. Similar to Figure 5.12a the activity patterns of these users display very pronounced bursts relatively to total traffic, which might be indicative of - thus a potential resource for the identification of - malicious activity (Niu et al., 2017).

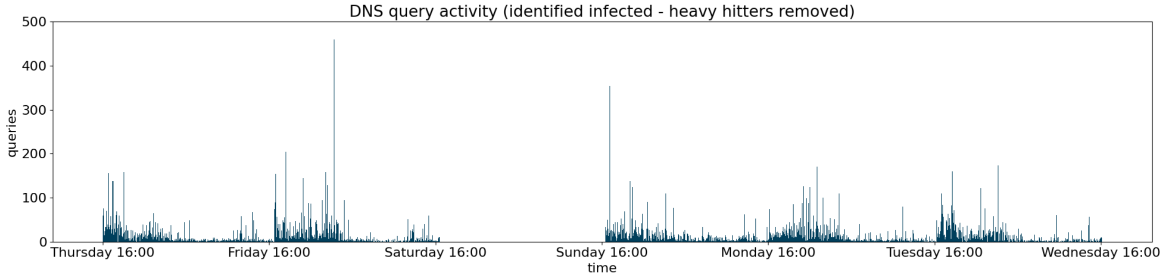
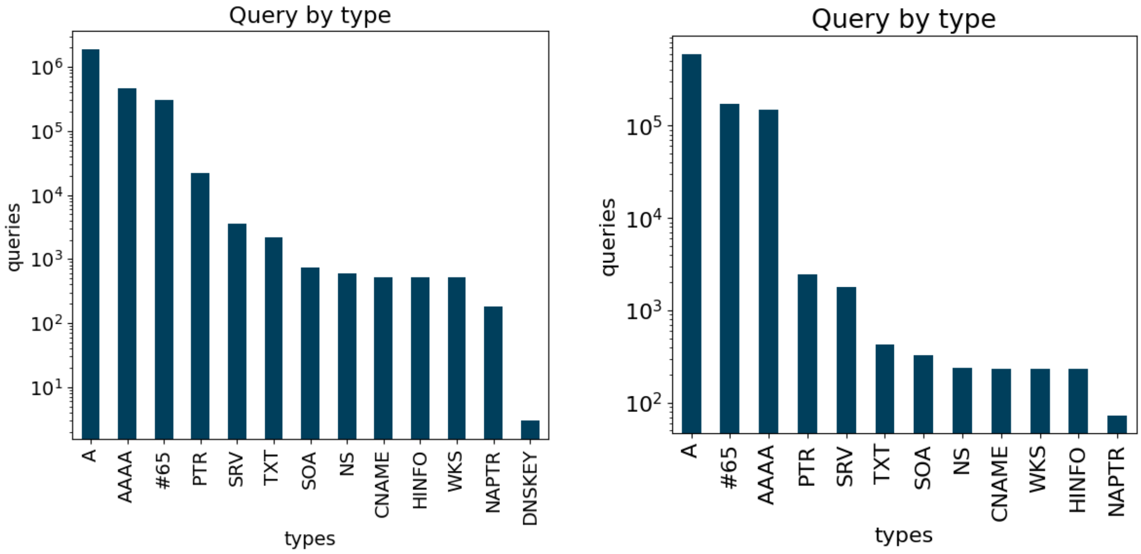


Figure 5.13: Temporal activity of 13 customers that made requests for malware-associated domains.

5.4.2. Request Characteristics

A number of features are extracted from the logged DNS queries in order to examine traffic characteristics such as the types of domains requested and features derived from the requested domain name. Figure 5.14 provides an overview of the most frequently observed response types among both the monitored users and the subset of interviewed users. Figure 5.15 provides an overview of the most commonly found top-level domains among the DNS requests among all monitored users and the interviewed users. Table 5.9 provides an overview of the most frequently requested domains among the dataset.



(a) Query return types among monitored users.

(b) Query return types among interviewed users.

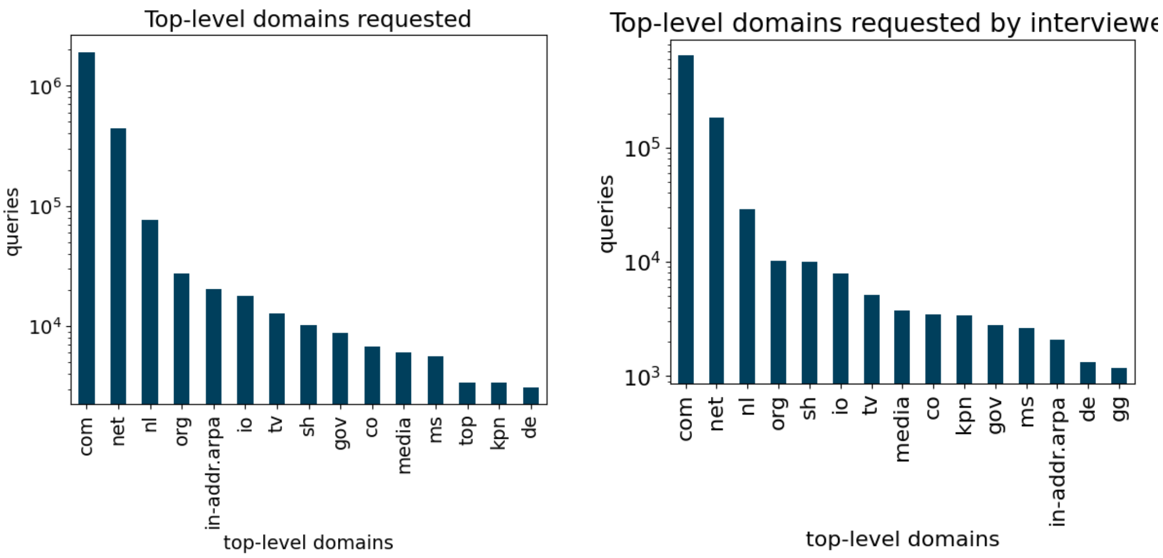
Figure 5.14: Most frequent query response types.

Top-Level Domains

The majority of DNS requests are aimed at the .com top-level domain; approximately 72% of all queries in the dataset. The .net domains account for a further 17% of all requests made; the vast majority of the remainder of the data set. The .nl and .org top-level domains account for a further 3% and 1% respectively, with the remainder of DNS requests directed at a variety of generic domains (such as .io, .tv, .cloud, and .media), at top-level domains associated with companies and organisations (.goog, .apple, and .kpn), and at (trans)national top-level domains (.de, .ru, .fr, .eu, and .be). One notable exception is the .arpa internet infrastructure domain (a largely deprecated TLD belonging to the United States Department of Advanced Research Projects Agency) whose in-addr.arpa domain is still commonly used to perform reverse DNS lookups.

Activity Types

Approximately 20% of the DNS queries are related to what could loosely be considered the ‘entertainment’ category. Notably, the popular video streaming service Netflix tops the chart of DNS activity, whose primary domain accounts for roughly 9% of all queried domain names by the monitored customers, and for roughly 11% of all



(a) Requested TLDs among monitored users.

(b) Requested TLDs among interviewed users.

Figure 5.15: Most requested top-level domains.

Table 5.9: The most requested domains among the monitored users as a percentage of all requests.

domain	**%**	**domain**	**%**	**domain**	**%**	**domain**	**%**
netflix	9.06	gstatic	1.53	kpn	0.94	nflximg	0.55
google	5.36	nflxso	1.52	akamai	0.81	honeywell	0.51
apple	4.46	doubleclick	1.45	root-servers	0.79	gvt2	0.50
googleapis	3.75	home	1.26	googlesyndication	0.77	youtube	0.49
tiktokcdn	2.98	icloud	1.25	aaplimg	0.73	live	0.48
facebook	2.41	tiktokv	1.24	dyndns	0.70	10	0.48
akamaiedge	2.38	akadns	1.18	fbcdn	0.66	dropboxapi	0.45
microsoft	2.16	apple-dns	1.14	amazonaws	0.62	googleusercontent	0.45
tp-link	1.86	synology	0.99	netgear	0.56	office	0.45
googlevideo	1.66	yahooapis	0.95	snapchat	0.55	spotify	0.45

queries when taking into account related domains used for content distribution (such as ‘nflxso’ and ‘nflximg’). To a lesser extent other (video) streaming services appear in the list of most requested domains, such as Google Video (1.86%) and YouTube (0.49%). Spotify (0.45%) is the only audio-streaming service with a notable but limited presence among the queries.

A second category of popular domains are related to news and information retrieval, which also account for approximately 20% of DNS requests. This category consists of domains such as google; the main domain for the popular search engine, as well as some of the company’s other solutions such as Google Drive. Similarly, the microsoft domain, which hosts a variety of Microsoft-related services such as its web-shop, product information pages, documentation pages and more, is a frequent occurrence among the DNS queries at approximately 2% of all requests made. Alongside Microsoft’s primary down, its outlook service (hosted at live.com), office and office365 domains, MSN services, and the Bing search engine. Apple’s apple and icloud domains, Amazon’s web services, and DropBox account for limited but notable fractions of DNS traffic.

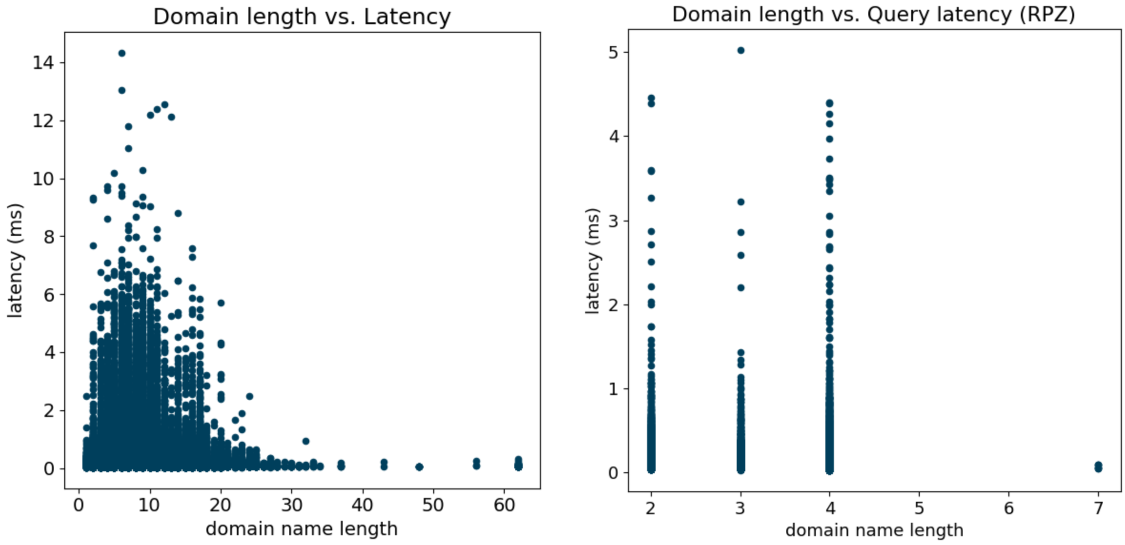
Social media sites and applications make up approximately 8% of the requested domain names. TikTok and its content distribution servers are most frequently encountered at approximately 4.5% of all DNS requests made. Facebook comes in at second place with roughly 2.5% of DNS requests inquiring about the Facebook domain name. Other popular social media websites and apps such as SnapChat, Instagram, and Whatsapp each account for roughly 0.5% of the observed DNS queries.

A further 4% of DNS requests are associated with requests to domain of suppliers of smart-home devices and networking hardware. 2% of these queries concern a request of the TP-Link domain which produces smart-home devices such as IP cameras, smart switch, and intelligent lighting, as well as routers, controllers, and other networking equipment. Another 1% of DNS requests are associated with Synology, which produces primarily network attached storage (NAS) devices. Lastly, approximately 0.6% of DNS queries are related to Netgear (networking equipment, NAS systems, and other smart-home devices) and Honeywell (climate control devices, including many smart-home heating solutions) both.

Latency & Domain Features

As delineated in the literature review in Chapter 2, a variety of features obtained from DNS queries may provide hints of malicious activity. Time-based features such as latency, and content-based features such as the length of requests or the number of consecutive alphabetical characters have previously been used in efforts to identify queries of malicious domains. While the collected data provides no opportunity for analysis of features such as TTL, the relationships between known or calculable features such as latency and content can be explored in an effort to identify remarkable patterns that might indicate abnormal activity.

Figures 5.16 visualizes the relationship between domain name length and latency. It can be observed that a large fraction of queries resolve with low latency, and that latency appears to be largely unrelated to the length of domain names. The vast majority of domain names requested have a length between 1 and 20 characters, with outliers of up to 62 characters of domain name length.



(a) Domain name length vs. latency

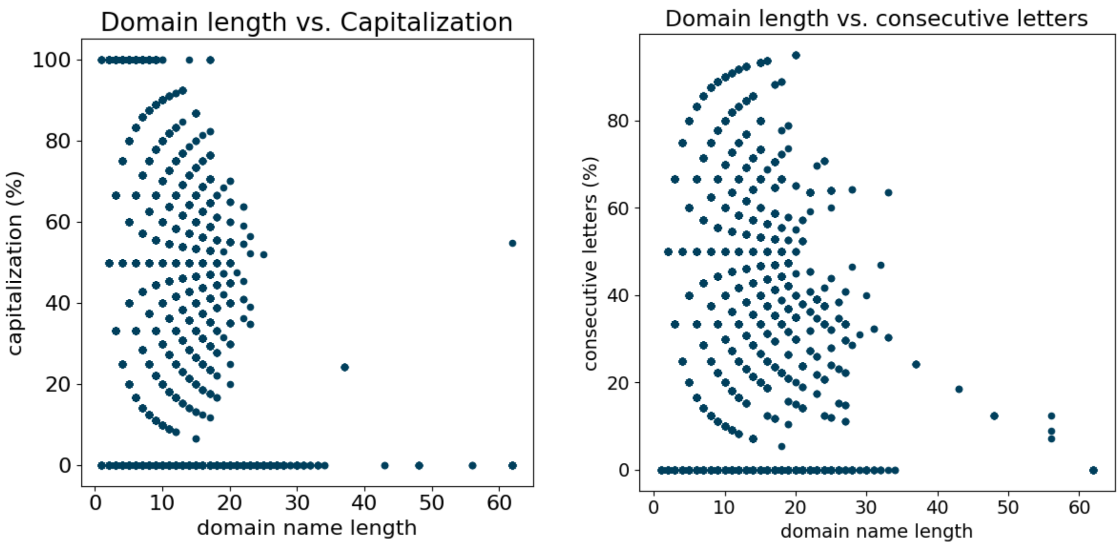
(b) Domain name length vs. latency (RPZ)

Figure 5.16: Domain name length plotted against query latency for RPZ and non-RPZ queries.

Figures 5.17a and 5.17b visualize the relationship between the length of the requested domain names and their capitalization and percentage of consecutive letters respectively. While figure 5.17a displays an interesting pattern in the percentage of capitalized letters in a domain name, these mix-case requests may simply represent a deliberate effort to increase entropy for purposes of spoofing resistance (known as 0x20 bit encoding). A similar pattern can be observed in figure 5.17b, where the percentage of consecutive letters relative to the length of the domain name is visualized. With the exception of a number of domains with excessively long names and a handful of domains toward the higher end of the latency spectrum, no significant outliers can be among the queries. Conventional statistical approaches using outlier detection may therefore not be highly valuable in the identification and elimination of malicious activity.

Table 5.10 lists the longest domain names encountered in the dataset which occurred more than once. It should be noted that the appearance of the domain name ‘10’ is an anomaly caused by the parsing of DNS requests into top-level domains, domain names, and subdomains. Inspection of queries which are categorised as relating to this domain name reveals that these are queries are in fact service discovery queries (i.e. queries of type lb._dns-sd._udp.<ip address>) and not reminiscent of malicious activity. Nevertheless, some interesting domain names can be identified:

- A number of requests involve domains and/or sub-domains that appear to be entirely randomly generated. Such high-entropy domain names are likely to be computer generated, a known weak point in blocklisting-based mitigation techniques (Tanaka et al., 2017), and may be indicative of illegitimate activity. Random domain name generation algorithms are commonly found in Fast Flux networks notorious for harbouring



(a) Capitalization as a percentage percentage of alphabetical characters in a domain name.

(b) Largest sub-string of consecutive letters as a percentage of domain name length.

Figure 5.17: Domain name length plotted against capitalization percentage and the largest fraction of consecutive letters in a domain name.

C&C infrastructure (Le Pochat et al., 2020), an evasion technique that is notably used by the Avalanche malware which was most frequently encountered among the infections identified in Section 5.3.

- A number of domains are composed of three consecutive but seemingly random words such as 'nationaldelinquencydelinquency'. In contrast with the mix-case domain names encountered before, the composition of these domain names may be the result of attempts to deliberately reduce the entropy of a request in order to avoid detection by entropy-related detection techniques. N-gram based methods such as those suggested by Liangboonprakong & Sornil (2013) and (Selvi et al., 2019) may be used to identify such domains, although it should be noted that legitimate domain names may follow a similar naming pattern.
- Two domain names include deceptive strings; the first to instagram.com-verify-<xyz>, which is a common method to verify domain ownership but in this case is an illegitimate link which has been blocklisted for phishing attempts, and the second domain 'riskfreeappinstalldeviceinstall.cyou'. Defending against threats like these might be significantly more difficult, as the employed domains more strongly resemble legitimate services and the relatively low number of requests may make such queries more comparable to those of regular requests.

5.4.3. Relation to the Research Interviews and Mitigation Efficacy

First and foremost, the characterisation of the DNS activity might explain some of the concerns voiced by interviewees with regards to the perceived drawbacks of the malwarefilter. Approximately 20% of all examined DNS traffic was related to productivity services, which might explain why a number of customers noted that the potential costs they associated with the service were related to the manner in which traffic blocking techniques might affect legitimate (but seemingly illegitimate) traffic. In line with the responses of interviewees, a near-constant stream of query activity was identified among users that noted ownership of a variety of IoT devices while among owners of traditional and mobile computing devices requests were made typically only during the afternoon and evening.

One interviewed user of the malwarefilter explicitly noted the use of a VPN service and the DNS activity recorded by this user consequently displayed only a handful of requests throughout the examined period. The use of such services may thus present a hurdle in the successful implementation of centralised DNS-based malware mitigation, as the mere use of DNS servers other than those provided by the ISP may severely impede the ability of such services to provide effective protection to end-users. In fact, the degree to which individuals are aware of the interplay between security measures and privacy-enhancing services may not only limit the degree of protection that service providers such as ISPs can ultimately provide, but may result in end-users unjustly considering their devices to be protected by a service that in fact fails to provide any protection at all. Thus, it is critical for the provider of centralised security measures to provide accessible and adequately detailed information regarding the functioning of their services and the manner in which their operation may be affected by other products. Transparency and information provision as a whole should be a focal point for these providers, in line with the themes

Table 5.10: Unusual domain names encountered among the longest domain names in the dataset.

Type	Query	Requests
Malware	accommodationinfractructuretwo.com.	13
	accommodationinfractructuretwo.com.	13
	exhaustedannulmentaccredited.com.	10
	nationaldeliquencydeliquency.com.	2
	bureaucracyambiguousfellow.com.	17
	disarrayanticipatedversion.com.	8
	salutationcheerlessdemote.com.	115
	www.notorietycheerypositively.com.	9
	classicalservicewaistcoat.com.	6
Scam	instagram.com-verify-9864254876345679154344343786875475575.info.	8
	www.riskfreeappinstalldeviceinstall.cyou.	4
Computer generated	53udy5-61mrmc2e2g5bzd5o0x54.i-y3op-sv7kht05asi64lt84457.com.	4
	cnk86p8f99tdhjy5pugouskebg.9pksm0sfyqzwqq8xo8m46nky041.com.	4
	jcgmqhuae836f4.qxlrptbs0u4k-l5q7659rmthtb5.com.	4
	z1rwwjt3bnzd0xw.2tp0jda7fml1625ta8zuvphv5tj.com.	4
	j8n7qr3u3-rkrrha.009yjjv1n4zc-n1swh-zz8ll70d.com.	2
	x-8x99z22irin3aa0d2o1.qbknj895q4omhnbzqgjnx6rai3y.com.	2
	yxmuo2pii3.1cipv-e22173jeck5iov3fz4wlr.com.	2
	2b763w7kiym0kz2sw4.swpc52t7qumz0fgz0pk2ht9bmb.com.	4
	79lfeamel61inqde.klj6-skbldcvms4soay-horuno.com.	4
	r07jlvbju.vrssdc-33-rv7fqrdnw5vnbuag.com.	4

identified in Section 5.1.

Based on the DNS queries logged by the malwarefilter a number of observations can be made about the potential value of this data in enhancing mitigative capabilities. First and foremost, a comparison between the temporal activity patterns of (1) the complete dataset, (2) queries modified by the response policy zone, and (3) a number of customers with a known malware infection revealed that among the latter two groups a number of notable activity spikes can be observed at several points in time. The sudden nature of these spikes - they are not preceded by a buildup of activity or followed by a gradual reduction in activity - might be an indication of the potentially illegitimate nature of these queries and could therefore prove to be a valuable resource to strengthen the ability to eliminate malicious traffic beyond the blocklist-only approach currently implementation of the malwarefilter.

Manual inspection of a number of queries that requested related to exceptionally long domain names revealed the existence of domain names which are known to be associated with malware (despite the fact that none of these queries were recognised as malware by the system) or phishing activities. These potentially malicious domains included both computer-generated names, which are likely associated with Fast Flux networks that harbour malware such as the Avalanche malware frequently identified among the monitored subjects in Section 5.3, and names that are seemingly random combinations of English words. While definitively establishing the nature of the requested domains and the capabilities of more intricate detection techniques is not in the scope of the study, the existence of such domain names among the examined DNS requests hints at the potential value of malware identification using temporal relations, n-grams, lexical analysis, or similar approaches based on the data recorded by the malwarefilter.

6 Synthesis

The final chapter of this report, chapter 6, presents the synthesis of the research effort. Section 6.1 discusses the findings embedded in the experimental results and the implications of these findings for technology, policy, and society. Section 6.2 lays out the limitations associated with the research effort and their importance in relation to the interpretation of results and the validity of the findings. Section 6.3 reiterates the most important findings and limitations, the implications they carry for policy and future academic efforts, and presents avenues for future research efforts.

6.1. Discussion

The study set to examine home-users' perceptions of online threats and the security of internet connected devices, their perceptions of - and willingness to adopt - centralised DNS-based malware mitigation as an effective way to deal with such threats, and the real-world efficacy of such measures in the mitigation of malicious activity. In line with the recommendations by Ifinedo (2012) and others, qualitative methods have been incorporated to study factors underlying the commonly used predictors for protection motivation, and to achieve a more comprehensive overview of motivations and barriers in the adoption of protective behaviour or tools. One often-cited deficiency in research concerning the adoption of security measures is the lack of evidence regarding the efficacy of the proposed tools or behaviour. This study has sought to incorporate quantitative evidence to assess not only the degree to which respondents actually adopt the adaptive coping behaviour, but also the extent to which it is effective in providing protection.

An investigation into security and threat perceptions among home-users of internet connected devices revealed that individuals predominantly consider internet-based threats to be a potential danger to themselves or the performance of their devices. Identity fraud and other forms of theft of sensitive data are the most frequently mentioned potential consequence of device compromise in line with earlier findings (Van Schaik et al., 2017; *Consumer Perception of Cyber Security Threats*, 2020). Notably, few end-users recognize the fact that compromised devices may be used by threat actors to harm individuals or organisations beyond the user or owner of a device. An inability to properly assess the danger posed by online threats as well as the degree to which devices were adequately secured was noted by a significant number of respondents. Reasons for this inability are found in an absence of prior experience with security incidents or an overall lack of knowledge about information technology and security, in line with research by Kulyk et al. (2020) who identify personal experience, media reports, and word of mouth as factors in security perceptions, and a plethora of earlier work that identify home-users as risks to themselves and others due to a lack of cyber awareness (S. Furnell et al., 2008; S. M. Furnell et al., 2007; Kritzinger & von Solms, 2010).

Despite these difficulties, and the doubts expressed by many customers about the adequacy of the default security measures afforded on internet connected devices, and the majority of end-users consider themselves ultimately responsible for their security. Other researchers have achieved similar results Thompson et al. (2017), and the findings of C. L. Anderson & Agarwal (2010) and others regarding the impact of psychological ownership on the willingness to adopt security measures further support these claims. The role of suppliers such as the device manufacturer or internet provider are mostly identified as supporting the user in maintaining the security and integrity of their internet connected devices through software updates and the enforcement of security habits, rather than the provision of ultimate 'secure' devices. The findings of Haney et al. support the idea that, while end-users hold manufacturers partially responsible for device security, they often doubt their willingness to spend extra money/time on improving the security of devices. Nevertheless, Haney et al. (2021) conclude that while concern and personal responsibility are often strong indicators of the willingness to engage in protective behaviour, there is a disconnect between willingness and capability among the queried users.

ISPs are in a unique position to relief much of the burden placed upon home-users to adequately secure their internet connected devices in light of the growing knowledge gap and complexity of cyber attacks Kritzinger & Von Solms (2013). The benefits of centralisation allow ISPs to implement measures for which individual users could never justify the cost and that could be managed and updated much more effectively. This change in responsibility of the ISP vis-a-vis the end-user brings with it legal and financial implications, and it may be argued that while ISPs can provide users with cyber-security information, they cannot ensure that they understand and are capable of implementing safeguarding measures (Kritzinger & Von Solms, 2012) as was also found in this study. Home-users were queried about their motivations for (non)adoption of a proposed centralised DNS-based malware mitigation service; the perceived benefits or drawbacks, and the role of the ISP.

Trust, coupled with the privacy implications of delegating security to a third party, is often cited as a dominant factor in the willingness to adopt centralised security measures. Trust in this sense implies trust in both the party providing the security measure, as well as parties it is affiliated with or subcontracts services from. Kulyk et al. identify a reliance on company reputation as an important element in perceptions of smart device security and privacy implications associated with using these devices. Yet, despite the prevalence of this theme among interviewees, statistical analysis yielded no evidence supporting trust as a major determinant in the use of centralised security measures. This apparent contradiction might imply that - while a certain threshold level of trust may be a prerequisite to considering the use of centralised security measures - it has little effect beyond enabling this consideration. Conversely, an explanation may be found in that trust in technology systems is often related to factors such as performance or functionality, helpfulness, and reliability which may be implicitly covered by other variables included in the conceptual model (Van der Werff et al., 2018).

Important barriers to the use of centralised security measures are perceived limitations in the efficacy of such services, as well the fact that these measures must bring some degree of added value (i.e. they compete with end-users' existing security setup). Many customers considered their own measures adequate in protecting their devices and thus saw no reason to enable the malwarefilter. The importance of end-users' perception of their own ability to provide adequate protection is supported by statistical evidence indicating that self-efficacy has a significant negative impact on the intention to use centralised security services. Earlier studies on home-user security behaviour find that self-efficacy is commonly a determinant in the use of decentralised security measures (Woon et al., 2005; Young et al., 2016; Hanus & Wu, 2016). Despite the fact that self-efficacy was often cited as a reason not to enable the malwarefilter, monitoring efforts proved that recurrence of infections was highly common among end-users that did not enable the mitigation service. Kovačević et al. (2020) find that 'self-identified experts' tend to exhibit less secure behaviour than self-identified non-experts. Martens et al. (2019) find in earlier research that an overestimation effect might be apparent among respondents whom consider themselves knowledgeable, capable, and aware of (malware) threats. This overestimation of users' own abilities, and the subsequent underestimation of the likelihood of victimization, is supported by the findings presented in this study as well as other works (West, 2008; M. Van Eeten & Bauer, 2009).

End-users primarily perceive the benefits of such centralised services to be related to ease of use or management, a feeling of safety, and as an additional (contingency) measure in addition to users' own security practices or services. The significance of both perceived response efficacy is supported almost universally among earlier works investigating the use of anti-malware services (Y. Lee & Larsen, 2009; Liang & Xue, 2010; Young et al., 2016; Martens et al., 2019), although Ophoff & Lakay (2018) notably do not find support for this in the context of ransomware. Explicit recommendations by an authoritative party such as the ISP may provide a further incentive to investigate the use - or outright enable - security measures as indicated by the significant positive impact of subjective norms on the intention to adopt centralised security measures. The significance of subjective norm is supported by the majority of previous works that involve these constructs or other factors relating to external pressure, such as social influence (Y. Lee & Larsen, 2009; Johnston & Warkentin, 2010; Ifinedo, 2012). While Thompson et al. (2017) found that subjective norm is a capable predictor neither in the context of desktop computers nor mobile devices, evidence is found for the significance of descriptive norm (perceptions of the behaviour performed by others, without the need for explicit social interactions) in predicting protective behaviour.

The efficacy of the solution was assessed through a comparison of infection cases and characteristics among users and non-users of the malwarefilter. While the number of malware infection decreased among both of those groups, a significantly much stronger reduction was observed among those that did enable the service. Simultaneously, a significant increase in the average duration of malware infections was found among both groups, while no significant difference was found between users and non-users of the malwarefilter. One possible explanation for the lack of effect on the duration of malware infections is the limited amount of feedback it provides to end-user regarding its functioning, and thus the need to take remediation actions is not communicated. Additionally, use of the malwarefilter was primarily common among owners of smart-home devices who might not at all be aware of the functioning or malfunction of their devices and who might not receive the warnings that are provided by the malwarefilter to those who, for example, navigate to a malicious website in their web browser. A lack of transparency with regards to the functioning of the malwarefilter, both in terms of being able to assess its value and in terms of receiving communications of the threats that it has mitigated, was one of the major deficiencies identified by end-users. This lack of transparency brings with it another issue; the fact that users might not be aware of the interactions between various services or products they use and the resulting (potentially diminished) effects on their value and mitigative capability. Notably, the use of VPNs and similar services may impact the malwarefilter's ability to provide meaningful protection at all.

Activity patterns for most users followed a schedule dictated by a regular working day, where the majority of requests are made starting in the afternoon and diminishing again late in the evening with small but noticeable spikes during the morning. Activity patterns for domains that were modified by an instated response policy zone or other policy reasons, indicating that the ISP has decided to redirect the queried domain for one reason or another, were found to largely follow the same temporal pattern as non-policed DNS traffic with the exception of a handful of notable activity spikes which may be indicative of behaviour that is not triggered by legitimate end-user activities. A visual exploration of query features such as domain name length, query latency, and the number of capitalized or consecutive letters in a domain name revealed no easily distinguishable subset that could be associated with malicious activity, although the examined features had previously been used to successfully identify malicious DNS queries in works which examined them with greater rigour, for example by subjecting them to machine learning techniques (Bilge et al., 2011; Shi et al., 2018). Nevertheless, a number of known-malicious domains were identified from among a set of queries with exceptionally long domain names. Detection techniques such as those suggested by J. Lee & Lee (2014), which leverages temporal relationships between queries, or those suggested by Tanaka et al. (2017) and Selvi et al. (2019) may be able to provide additional mitigative capability but their application could not be investigated within the bounds of this study.

6.2. Limitations

IoT-devices and Ownership

The predominant limitations of the study are associated with the sampling methods and the availability of a subject set that represents the broader base of home-users of computing devices, specifically smart-home devices or other internet-of-things applications. The research set out to examine the adoption of centralised security measures by end-users and their efficacy in protecting owners that might otherwise have been unaware of a malware infection or unable to protect their devices through traditional (distributed) services or software such as anti-virus software. These specific users represent a limited fraction of the final sample of responses in both the research interviews (a third of participants) and the questionnaire (approximately half of respondents), and the limited prevalence of IoT-based malware infections among the analysed data. The study may therefore have failed to capture the beliefs of this group of end-users to the extent that it set out to do, and is restricted in its assessment of the benefits derived from centralised DNS-based services in mitigating IoT-based malware.

Response and Non-response Bias

In both the cases of the research interviews and the questionnaire used to assess the conceptual model, one should take into account that the responses received may be biased both in their representativeness of the overall population from which the samples were drawn and the responses these subjects provided.

Voluntariness bias may affect the representation of certain groups of end-users in the study, as appears to be the case due to the skewed distribution of respondents toward highly-educated, middle-aged men in both the research interviews and the questionnaire responses. It is not possible to definitively determine to what extend this skewedness is a product of the distribution from which the sample was drawn, whether it may be related to the demographics of the KPN customer base itself, or whether it is a product of the sampling method. However, earlier research efforts at the KPN Abuse Desk have achieved similar sample demographics (Altena, 2018; Verstegen, 2019; Bouwmeester, 2020), and the obtained demographics exhibit some degree of similarity to demographic provided by the ISP.

Additionally, limitations on the sample that has been obtained and the related data may have been imposed by the degree to which contacted customers have received and read the malwarefilter notification, and its clarity and comprehensibility. In order to ensure that the notification adequately conveys the intent of the study, it has been constructed in accordance with guidelines established in earlier research and checked and approved by the Abuse Desk. However, little can be done to ensure that the notification reaches the correct customers or persons within a household. The notification is sent to the email address known to KPN to be associated with the IP address of the customer. Similarly, phone number associated with customer profiles may be outdated, or customers may be unreachable for other reasons.

Furthermore, social desirability bias may be encountered among both the research interviews and questionnaire. This is further aggravated by the fact that customers were contacted by the researcher in the position of an employee of the internet service provider. The internet service provider may be considered an authoritative party by the contacted customers, especially in the context of device security incidents, and they may therefore be even more likely to provide responses that are perceived to be desirable.

Evolution of the Threatscape

With regards to the assessment of the efficacy of the malwarefilter beyond the limited inclusion of IoT-device owners, a second limitation must be noted in the dependence of the assessment on developments within the threatscape and the wider cyber security environment. This includes both the actions of threat actors and the behaviour of end-users (and anyone in between). Threat actors may affect the amount and characteristics of observed malicious activity directly by introducing new threats, iterating on existing threats, retiring known threats, and more. End-users may alter their behaviour in response to media attention, communications sent by the ISP, or for other reasons. Security service providers continually adapt their products to deal with new threats and actors such as the ShadowServer foundation on which the ISP relies for much of its identification of compromised hosts may be successful or unsuccessful in identifying threats at changing intervals. The actions and behaviour of all of these actors affect the assessment of malwarefilter, as well as the actions and behaviour of other actors. As such, one should be careful to consider any of the findings as a definitive answer rather than an evaluation at a single point in time.

Similarly, the qualitative data that has been collected in the form of research interviews may have been affected by developments surrounding the cyber security environment at the time the interviews were conducted. Notably, prior and during the period in which the research interviews were conducted a number of news reports were released in a national newspaper on the relationship between the internet service provider and a foreign supplier of part of its core infrastructure.

Malware Infection Data

Both in the cases of the malware infection logs compiled by the Abuse Desk and the DNS logs generated by the malwarefilter service one should account for a degree of incompleteness in the data or other anomalies that may be associated with the collection of these events.

The malware infection logs are likely to be incomplete in the sense that it cannot be expected to all actual infections that have taken place within the examined periods; notably it may have failed to capture novel or especially intricate malwares that are capable of avoiding detection. Furthermore, most of the infections are reported by third-party reporters such as the ShadowServer foundation, which may have experienced issues or difficulties in the provision of reports at certain points throughout the monitoring periods and therefore provided an incomplete overview of known infections with the ISP's autonomous systems.

Additionally, the fact that the interviews are conducted during the monitoring period may influence the reliability of the infection case data. The contacted customers might be inclined to more actively engage in protecting their devices or the search and elimination of existing infections after having been called by the researcher in regards an interview.

DNS Query Data

The DNS logs are likely to be incomplete primarily due to limitations associated with the platform that is used to log malwarefilter users' DNS activity and (business) decisions to support certain functionalities. Notably, although the logs provide a systematic overview of the DNS activity of users subscribed to the malwarefilter, it was not originally set up or intended for purposes of systematic monitoring.

At the time the DNS log data was gathered the platform recorded approximately 30 hours of the most recent requests made. The DNS data is extracted on a 24-hourly basis, but this may nevertheless prove to result in an incomplete dataset. Additionally, the number of records that can be extracted at once is limited to 30,000 which results in data loss if a single IP has made more than 30,000 DNS queries within the 24-hour period. Gathering a truly complete dataset within the limitations of the platform would result in an excessive amount of manual labour (if at all possible) and is therefore not considered to be feasible.

6.3. Conclusion

This study has sought to examine the adoption and efficacy of centralised malware mitigation measures by end-users through an investigation at one The Netherlands' largest internet service providers. A literature review was conducted to identify established theories on the adoption of information security measures and malware mitigation measures by end-users specifically. Based on these theories, a conceptual model to research the adoption of centralised malware mitigation measures was developed, and an experiment conducted to explore customers' willingness to adopt the ISP's DNS-based malwarefilter service.

"What are the main concerns in using centralised DNS-based malware mitigation services, how effective are such services at reducing malicious activity, and how does the DNS activity of legitimate users compare to that of compromised devices?"

Most customers employ software solutions and regularly perform actions such as security updates to keep their devices safe, but ultimately find it difficult to assess the degree to which they are adequately protected due to a lack of understanding of both threats and countermeasures. This appears to result in an overestimation of one's capability to secure devices, as evidenced by the recurrence of infections among especially non-users of the malwarefilter. While trust in both the ISP itself as well its suppliers or subcontractors is noted by many users to be instrumental in the consideration to enable the malwarefilter, no quantitative evidence was found to support this as a significant factor in the intention to adopt such services.

With regards to centralised malware mitigation specifically, end-users are primarily worried about the impact of on their productivity; the risk of legitimate or purposely generated traffic being eliminated by a service they have little or no control over. While financial costs are also semi-frequently mentioned as a potential drawback of security services provided by an ISP, perceived costs were not found to be a significant determinant in predicting the use of such measures. Additionally, customers may consider there to be little added value to such services in addition to the (distributed) measures they employ themselves or consider it at all unlikely to be effective at mitigating malicious activity. Conversely, the use of such measures in addition to one's own (distributed) security is considered by many as one of the main reasons to adopt such services; to provide an extra layer of protection. Quantitative evidence supports perceived response efficacy and self-efficacy as significant determinants for the intention to adopt centralised security measures.

The malwarefilter was found have a significant effect in limiting customers' exposure to malware, sharply reducing the number of infections incurred by customers that enabled it relative to the group of customers that decided not to enable the service. Use of the malwarefilter did not affect the average duration of malware infections significantly, possibly due to a lack of visibility of the mitigating actions it takes which was commonly reported as a drawback of the service. Providing greater insight into not only the functioning of centralised security measures but also into the actions taken to protect devices and their users may aid the adoption of such measures, while simultaneously increasing users' awareness of online threats and ensuring the efficacy of the mitigation in light of potential interference with other services.

The DNS activity of malwarefilter users indicates that households that enabled the service predominantly use their internet connection for productivity and entertainment purposes, potentially explaining some customers' concerns about the malwarefilter's impact on productivity. Notable spikes in temporal activity were observed among both users with a known or suspected malware infection as well as among the queries that were modified in line with the response policy zone, which might be indicative of illegitimate DNS activity. An examination of query characteristics and derived features did not hint at their significant value in the identification of malicious traffic. Manual inspection of a number of outlier requests based on domain name length revealed the existence of several known-malicious or otherwise suspicious domain names which could potentially be identified using a variety of methods proposed in other works, although the merits of these methods could not be explored within the bounds of the research.

6.3.1. Implications & Recommendations

Academic Implications

Martens et al. (2019) identified discrepancies among the applicability and significance of various PMT constructs in the context of different types of malicious cyber activity. This study finds that, in line with these and other earlier works, the significance of several PMT constructs and their counterparts in theories such as the health belief model and technology threat avoidance theory are often dependent not only on the threat-context, but also on the characteristics of the prescribed tool or behaviour and the devices being secured (Thompson et al., 2017).

While self-efficacy is almost universally found to be a significant predictor in IS research, the manner in which this construct affects the adoption security measures or behaviour should be examined in greater detail with respect to the exact technology or behaviour that is considered as the adaptive coping response and how this differs between various technologies (both in terms of safeguarding measures, and in terms of the vulnerable devices). Although this study did not yield enough responses to build individual regression models for each of the types of end-users with regards to device ownership, this may prove an interesting effort for works with a wider reach and to reexamine the applicability and power of these models in consideration of new environments and technologies.

Inclusion of the construct of maladaptive coping with emotions provided insights that are in line with the few earlier works that have examined maladaptive coping methods. The data collected through the research interviews supports the idea that such emotional coping methods are at least somewhat prevalent when it comes to cyber security - several interviewees expressed ideas such as the inevitability of online threats or denied potential harm to their privacy - but the exact manner in which this affects end-users' cyber security behaviour was not examined in this study.

Unlike most earlier studies on the adoption of information security measures or behaviour, this study attempted to recruit a significantly diverse set of subjects, beyond the college or university student populations that are often used as a convenience sample in earlier works, by recruiting subjects from a large national internet service provider. Despite these efforts, the samples for both the research interviews and questionnaire display significant skewedness towards a highly-educated, middle-aged, predominantly male population. As such, future efforts should once again aim to diversify the dataset or subject populations potentially through other sampling means than voluntary participation, or by offering rewards for participants from underrepresented groups.

Abuse Handling

The findings also have implications for the policies maintained by internet service providers such as KPN in maintaining and aiding the security interests of their customers. For ISPs, one of the main consideration should be to re-evaluate the manner in which customers are contacted with regards to security incidents. The most practical findings of the study indicate three important things:

Firstly, a significant fraction of customers is incapable of properly assessing the situation surrounding the security of their internet connected devices, and many vulnerabilities appear to remain after customers have previously suffered from a malware infection or other malicious activity. While the reduction in infections among both groups of customers that received a notification of a malware infection relative to those that did not indicates some improvement in end-user security standards in result to such messages, the still relatively high degree of infection recurrence among customers that received such a notification but did not enable the malwarefilter may be considered as evidence that this elicited improvement is not sufficient.

Secondly, the findings indicate that the ISP and other authoritative parties may effectively encourage the use security measures by end-users by providing greater support and insights with regards to the dangers posed by online threats and the functioning and efficacy of security measures. The data gathered in this study shows that malware infections are highly likely to recur, even (or perhaps especially) among those who consider themselves capable of adequately protecting their devices. While customers might not be inclined to change their security habits or be capable of adequately securing their devices themselves, centralised measures such as the malwarefilter provide an effective means to shield these vulnerable customers and devices without the need for significant investment - financial, time, effort, or knowledge acquisition - on the side of the end-user.

Thirdly, opportunities to improve the abuse resolution process as employed by the KPN Abuse Desk. Most strikingly, the use of quarantine measures and walled gardens might force customers to resolve active infections without necessarily reducing the vulnerability of their devices to future malware infections. As such, explicitly recommending the use of the malwarefilter either in addition to current remediation protocols or as a replacement thereof might provide an effective way to reduce abuse cases. Specifically, the malwarefilter might be an effective manner to reduce malware infections among repeat offenders and its use as an alternative to high-impact measures such as quarantines should be investigated in cases where the use of these high-impact measures is not feasible, such as in the business market.

Position of the ISP

Despite the potential of the malwarefilter as a more widely integrated security service, account should be taken of the implications it might have for the transfer of responsibility perceptions from the customer into the hands of the ISP. While this study shows that customers almost universally look to themselves to secure their devices, regardless of whether they use the services offered by their ISP, there is a risk that end-users that enable the malwarefilter become complacent under the assumption that they have delegated device security to their ISP in full. If the malwarefilter is indeed integrated more closely in the security offer of KPN and incorporated in the abuse process, it must be made explicit to customers that their individual behaviour and the degree to which they employ additional security measures remains relevant in providing optimal or even sufficient protection against malware threats and other malicious activity.

The provision of information that is both accessible and sufficiently comprehensive may perhaps be considered the most important factor in both motivating the use of measures like the malwarefilter, as well as ensuring that these measure can operate to their full effectiveness. Interference between privacy-improving services such as VPNs and security measures that rely on the analysis of (DNS) traffic of which users might not be aware may give them a false sense of security which in turn may make these users more vulnerable to online threats rather than less vulnerable.

Future Work

Future work should aim to capture responses from the general population to investigate the degree to which the findings generalize beyond end-users with previous malware infections. While this study has been able to investigate the adoption and efficacy of centralised security measures among subjects whom had previously been vulnerable to malware, it cannot conclude to what extent the findings are affected by the predispositions of the sampled population with respect to the general population.

Additionally, future research might re-evaluate the model and its applicability to the delegation of device security, and specifically to re-evaluate the applicability of concepts such as threat appraisal and constructs such as response cost, which show inconsistent results both among earlier works and this study. The inclusion of threat awareness factors which are proven to be significant predictors for threat appraisal and often cited as a reason for lacking significance of threat appraisal in predicting protective behaviour should be studied further. The construct of response cost needs to be investigated further to study the impact of non-financial components of perceived costs, such as the elimination of legitimate traffic or other forms of productivity loss which were found to be often cited perceived drawbacks of (centralised) DNS-based security services in this study.

Lastly, future research on the adoption of centralised security measures should consider including direct measurements of the impact of recommendations or security notification sent by authoritative parties, or extend the model or the models it is based on in novel directions based on the presented findings. The demonstrated significance of subjective norm, alongside the emergence of authoritative communications as an important theme in the consideration of such measures by end-users, provides an avenue to study the impact of social factors in the decision to delegate the internet security of household appliances to a third party.

References

Abhishta, A. (2019). *The blind man and the elephant: Measuring economic impacts of ddos attacks*. University of Twente.

Ahlmeyer, M., & Chircu, A. M. (2016). Securing the internet of things: A review. *Issues in information Systems*, *17*(4).

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179–211.

Alieyan, K., ALmomani, A., Manasrah, A., & Kadhum, M. M. (2017). A survey of botnet detection based on dns. *Neural Computing and Applications*, *28*(7), 1541–1558.

Allix, K., Jérome, Q., Bissyandé, T. F., Klein, J., State, R., & Le Traon, Y. (2014). A forensic analysis of android malware–how is malware written and how it could be detected? In *2014 ieee 38th annual computer software and applications conference* (pp. 384–393).

Altena, L. (2018). Exploring effective notification mechanisms for infected iot devices. *TU Delft Education Repository*.

Anderson, C. L., & Agarwal, R. (2010). Practicing safe computing: A multimethod empirical examination of home computer user security behavioral intentions. *MIS quarterly*, 613–643.

Anderson, R., Barton, C., Böhme, R., Clayton, R., Van Eeten, M. J., Levi, M., … Savage, S. (2013). Measuring the cost of cybercrime. In *The economics of information security and privacy* (pp. 265–300). Springer.

Antonakakis, M., Dagon, D., Luo, X., Perdisci, R., Lee, W., & Bellmor, J. (2010). A centralized monitoring infrastructure for improving dns security. In *International workshop on recent advances in intrusion detection* (pp. 18–37).

Antonakakis, M., Perdisci, R., Dagon, D., Lee, W., & Feamster, N. (2010). Building a dynamic reputation system for dns. In *Usenix security symposium* (pp. 273–290).

Arachchilage, N. A. G., & Love, S. (2014). Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, *38*, 304–312.

Asghari, H., van Eeten, M. J., & Bauer, J. M. (2015). Economics of fighting botnets: Lessons from a decade of mitigation. *IEEE Security & Privacy*, *13*(5), 16–23.

Atlam, H. F., & Wills, G. B. (2020). Iot security, privacy, safety and ethics. In M. Farsi, A. Daneshkhah, A. Hosseinian-Far, & H. Jahankhani (Eds.), *Digital twin technologies and smart cities* (pp. 123–149). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-18732-3_8` doi: 10.1007/978-3-030-18732-3_8

Bertino, E., Choo, K.-K. R., Georgakopolous, D., & Nepal, S. (2016). *Internet of things (iot) smart and secure service delivery.* ACM New York, NY, USA.

Bertino, E., & Islam, N. (2017). Botnets and internet of things security. *Computer*, *50*(2), 76–79.

Bilge, L., Kirda, E., Kruegel, C., & Balduzzi, M. (2011). Exposure: Finding malicious domains using passive dns analysis. In *Ndss* (pp. 1–17).

Boerman, N., Henneke, M., Moura, G., Schaapman, G., & de Weerdt, O. (2018). *The impact of ddos attacks on dutch enterprises.* Retrieved from `https://www.nbip.nl/wp-content/uploads/2018/11/NBIP-SIDN-DDoS-impact-report.pdf`

Bouwmeester, B. J. (2020). A visit to the crime scene: Monitoring end-users during the remediation process of mirai infected internet of things devices. *TU Delft Education Repository*.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative research in sport, exercise and health*, *13*(2), 201–216.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS quarterly*, 523–548.

Chiou, J.-S. (2004). The antecedents of consumers' loyalty toward internet service providers. *Information & Management*, *41*(6), 685–695.

Claar, C. L. (2011). *The adoption of computer security: An analysis of home personal computer user behavior using the health belief model* (Unpublished doctoral dissertation). Utah State University.

Claar, C. L., & Johnson, J. (2012). Analyzing home pc security adoption behavior. *Journal of Computer Information Systems*, *52*(4), 20–29.

Connery, H. M. (2013). *Dns response policy zones history, overview, usage and research.* Denmark: University of Denmark.

Consumer perception of cyber security threats. (2020, Jul). Retrieved from `https://blumbergcapital.com/cybersecurity2020/`

Cook, S. (2020, Nov). *Ddos attack statistics and facts for 2018-2020.* Retrieved from `https://www.comparitech.com/blog/information-security/ddos-statistics-facts/`

Crossler, R., & Bélanger, F. (2014). An extended perspective on individual security behaviors: Protection motivation theory and a unified security practices (usp) instrument. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *45*(4), 51–71.

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Deng, Z., Lu, Y., Wei, K. K., & Zhang, J. (2010). Understanding customer satisfaction and loyalty: An empirical study of mobile instant messages in china. *International journal of information management*, *30*(4), 289–300.

Dickson, B. (2019, Oct). *20 years of ddos attacks: What has changed?* Retrieved from `https://portswigger.net/daily-swig/20-years-of-ddos-attacks-what-has-changed`

Dinev, T., & Hu, Q. (2007). The centrality of awareness in the formation of user behavioral intention toward protective information technologies. *Journal of the Association for Information Systems*, *8*(7), 23.

Dodel, M., & Mesch, G. (2017). Cyber-victimization preventive behavior: A health belief model approach. *Computers in Human behavior*, *68*, 359–367.

Driscoll, D. L., Appiah-Yeboah, A., Salib, P., & Rupert, D. J. (2007). Merging qualitative and quantitative data in mixed methods research: How to and why not. *Ecological and Environmental Anthropology*.

Feily, M., Shahrestani, A., & Ramadass, S. (2009). A survey of botnet and botnet detection. In *2009 third international conference on emerging security information, systems and technologies* (pp. 268–273).

Feinstein, L., Schnackenberg, D., Balupari, R., & Kindred, D. (2003). Statistical approaches to ddos attack detection and response. In *Proceedings darpa information survivability conference and exposition* (Vol. 1, pp. 303–314).

Fugard, A. J., & Potts, H. W. (2015). Supporting thinking on sample sizes for thematic analyses: a quantitative tool. *International Journal of Social Research Methodology*, *18*(6), 669–684.

Furnell, S., Tsaganidi, V., & Phippen, A. (2008). Security beliefs and barriers for novice internet users. *Computers & Security*, *27*(7-8), 235–240.

Furnell, S. M., Bryant, P., & Phippen, A. D. (2007). Assessing the security perceptions of personal internet users. *Computers & Security*, *26*(5), 410–417.

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC medical research methodology*, *13*(1), 1–8.

Goth, G. (2007). The politics of ddos attacks. *IEEE Distributed Systems Online*, *8*(8), 3–3.

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PLoS One*, *15*(5), e0232076.

Haag, S., Siponen, M., & Liu, F. (2021). Protection motivation theory in information systems security research: A review of the past and a road map for the future. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *52*(2), 25–67.

Haney, J., Acar, Y., & Furman, S. (2021). " it's the company, the government, you and i": User perceptions of responsibility for smart home privacy and security. In *30th {USENIX} security symposium ({USENIX} security 21).*

Hanus, B., & Wu, Y. □. (2016). Impact of users' security awareness on desktop security behavior: A protection motivation theory perspective. *Information Systems Management*, *33*(1), 2–16.

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems*, *18*(2), 106–125.

Hinchliffe, A. (2019, Mar). *Dns tunneling: how dns can be (ab)used by malicious actors.* Retrieved from https://unit42.paloaltonetworks.com/dns-tunneling-how-dns-can-be-abused-by-malicious-actors/

Huang, Y., Geng, X., & Whinston, A. B. (2007). Defeating ddos attacks by fixing the incentive chain. *ACM Transactions on internet Technology (TOIT)*, *7*(1), 5–es.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, *9*(3), 90–95.

Ifinedo, P. (2012). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*, *31*(1), 83–95.

Institute, T. D. (2021). *Dnssec guide: What's edns all about (and why should i care)?* Retrieved from https://dnsinstitute.com/documentation/dnssec-guide/ch03s05.html

Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: tips for students new to the field of qualitative research. *Qualitative Report*, *17*, 6.

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: An empirical study. *MIS quarterly*, 549–566.

Jones, C. M., McCarthy, R. V., Halawi, L., & Mujtaba, B. (2010). Utilizing the technology acceptance model to assess the employee adoption of information systems security measures. *Issues in Information Systems*, *11*(1), 9.

Kara, I. (2019). A basic malware analysis method. *Computer Fraud & Security*, *2019*(6), 11–19.

Kaspersky. (2018, Feb). *Ddos breach costs rise to over $2m for enterprises finds kaspersky lab report.* Kaspersky Labs. Retrieved from https://usa.kaspersky.com/about/press-releases/2018_ddos-breach-costs-rise-to-over-2m-for-enterprises-finds-kaspersky-lab-report

Kim, T. H., & Reeves, D. (2020). A survey of domain name system vulnerabilities and attacks. *Journal of Surveillance, Security and Safety*, *1*(1), 34–60.

Kolias, C., Kambourakis, G., Stavrou, A., & Voas, J. (2017). Ddos in the iot: Mirai and other botnets. *Computer*, *50*(7), 80–84.

Kovačević, A., Putnik, N., & Tošković, O. (2020). Factors related to cyber security behavior. *IEEE Access*, *8*, 125140–125148.

KPN. (2020). personal communication.

Kritzinger, E., & Von Solms, S. (2012). A framework for cyber security in africa. *Journal of Information Assurance & Cybersecurity*, *2012*, 1.

Kritzinger, E., & Von Solms, S. (2013). Home user security-from thick security-oriented home users to thin security-oriented home users. In *2013 science and information conference* (pp. 340–345).

Kritzinger, E., & von Solms, S. H. (2010). Cyber security for home users: A new way of protection through awareness enforcement. *Computers & Security*, *29*(8), 840–847.

Kulyk, O., Milanovic, K., & Pitt, J. (2020). Does my smart device provider care about my privacy? investigating trust factors and user attitudes in iot systems. In *Proceedings of the 11th nordic conference on human-computer interaction: Shaping experiences, shaping society* (pp. 1–12).

Kumar, N., Mohan, K., & Holowczak, R. (2008). Locking the door but leaving the computer vulnerable: Factors inhibiting home users' adoption of software firewalls. *Decision Support Systems*, *46*(1), 254–264.

Kupreev, O., Badovskaya, E., & Gutnikov, A. (2020, May). *Ddos attacks in q1 2020.* Retrieved from https://securelist.com/ddos-attacks-in-q1-2020/96837/

Lafrance, A. (2016). *How much will today's internet outage cost?* Retrieved from https://www.theatlantic.com/technology/archive/2016/10/a-lot/505025/

Lau, F., Rubin, S. H., Smith, M. H., & Trajkovic, L. (2000). Distributed denial of service attacks. In *Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics.'cybernetics evolving to systems, humans, organizations, and their complex interactions'(cat. no. 0* (Vol. 3, pp. 2275–2280).

Law, K., Lui, J. C., & Yau, D. K. (2002). You can run, but you can't hide: an effective methodology to traceback ddos attackers. In *Proceedings. 10th ieee international symposium on modeling, analysis and simulation of computer and telecommunications systems* (pp. 433–440).

Lee, J., & Lee, H. (2014). Gmad: Graph-based malware activity detection by dns traffic analysis. *Computer Communications*, *49*, 33–47.

Lee, K., Kim, J., Kwon, K. H., Han, Y., & Kim, S. (2008). Ddos attack detection method using cluster analysis. *Expert systems with applications*, *34*(3), 1659–1665.

Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for information systems*, *12*(1), 50.

Lee, Y., & Larsen, K. R. (2009). Threat or coping appraisal: determinants of smb executives' decision to adopt anti-malware software. *European Journal of Information Systems*, *18*(2), 177–187.

Le Pochat, V., Maroofi, S., Van Goethem, T., Preuveneers, D., Duda, A., Joosen, W., … others (2020). A practical approach for taking down avalanche botnets under real-world constraints. In *Proceedings of the 27th annual network and distributed system security symposium.*

Li, W., Jin, J., & Lee, J.-H. (2019). Analysis of botnet domain names for iot cybersecurity. *IEEE Access*, *7*, 94658–94665.

Li, Y., & Siponen, M. T. (2011). A call for research on home users' information security behaviour. In *Pacis* (p. 112).

Li, Z., Liao, Q., & Striegel, A. (2009). Botnet economics: Uncertainty matters. In M. E. Johnson (Ed.), *Managing information risk and the economics of security* (pp. 245–267). Boston, MA: Springer US. Retrieved from `https://doi.org/10.1007/978-0-387-09762-6_12` doi: 10.1007/978-0-387-09762-6_12

Liang, H., & Xue, Y. (2009). Avoidance of information technology threats: A theoretical perspective. *MIS quarterly*, 71–90.

Liang, H., & Xue, Y. L. (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective. *Journal of the association for information systems*, *11*(7), 1.

Liangboonprakong, C., & Sornil, O. (2013). Classification of malware families based on n-grams sequential pattern features. In *2013 ieee 8th conference on industrial electronics and applications (iciea)* (pp. 777–782).

Libfeld, R. (n.d.). *What is dns spoofing man in the middle attack?: Security wiki.* Retrieved from `https://doubleoctopus.com/security-wiki/threats-and-tools/dns-spoofing/`

Ligh, M. H., Adair, S., Hartstein, B., & Richard, M. (2011). *Malware analyst's cookbook and dvd: tools and techniques for fighting malicious code*. Wiley Pub., Incorporated.

Ma, X., Zhang, J., Li, Z., Li, J., Tao, J., Guan, X., … Towsley, D. (2015). Accurate dns query characteristics estimation via active probing. *Journal of Network and Computer Applications*, *47*, 72–84.

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of experimental social psychology*, *19*(5), 469–479.

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society*, *14*(1), 81–95.

Martens, M., De Wolf, R., & De Marez, L. (2019). Investigating and comparing the predictors of the intention towards taking security measures against malware, scams and cybercrime in general. *Computers in Human Behavior*, *92*, 139–150.

McKinney, W. (2011). pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, *14*(9).

Menard, P., Bott, G. J., & Crossler, R. E. (2017). User motivations in protecting information security: Protection motivation theory versus self-determination theory. *Journal of Management Information Systems*, *34*(4), 1203–1230.

Mills, A., & Sahi, N. (2019). An empirical study of home user intentions towards computer security. In *Proceedings of the 52nd hawaii international conference on system sciences.*

NCSC. (2019, Oct). *Wees voorbereid op dot en doh.* Nationaal Cyber Security Centrum. Retrieved from `https://www.ncsc.nl/actueel/nieuws/2019/oktober/2/wees-voorbereid-op-dot-en-doh`

Ng, B.-Y., Kankanhalli, A., & Xu, Y. C. (2009). Studying users' computer security behavior: A health belief perspective. *Decision Support Systems*, *46*(4), 815–825.

Ngo, Q.-D., Nguyen, H.-T., Nguyen, L.-C., & Nguyen, D.-H. (2020). A survey of iot malware and detection methods based on static features. *ICT Express*.

Niu, W., Zhang, X., Yang, G., Zhu, J., & Ren, Z. (2017). Identifying apt malware domain based on mobile dns logging. *Mathematical Problems in Engineering*, *2017*.

.nl statistieken: Sidn labs. (2021, Jul). Retrieved from `https://stats.sidnlabs.nl/nl/dnssec.html`

Ophoff, J., & Lakay, M. (2018). Mitigating the ransomware threat: a protection motivation theory approach. In *International information security conference* (pp. 163–175).

Otgonbold, T. (2014). Adapt: An anonymous, distributed, and active probing-based technique for detecting malicious fast-flux domains. *Iowa State University Digital Repository*.

Pieterse, H., & Olivier, M. S. (2012). Android botnets on the rise: Trends and characteristics. In *2012 information security for south africa* (pp. 1–5).

Ramachandran, A., Feamster, N., & Dagon, D. (2006). Revealing botnet membership using dnsbl counter-intelligence. *Sruti*, *6*, 49–54.

Richards, A., & Smith, H. (2007). Addressing the role of private security companies within security sector reform programmes. *Journal of Security Sector Management*, *5*(1), 1–14.

Rowe, B., Reeves, D., & Gallaher, M. (2011). *The role of internet service providers in cyber security*. Citeseer.

Sachdeva, M., Singh, G., Kumar, K., & Singh, K. (2010). Ddos incidents and their impact: A review. *Int. Arab J. Inf. Technol.*, *7*(1), 14–20.

Safa, N. S., Von Solms, R., & Furnell, S. (2016). Information security policy compliance model in organizations. *computers & security*, *56*, 70–82.

Selvi, J., Rodríguez, R. J., & Soria-Olivas, E. (2019). Detection of algorithmically generated malicious domain names using masked n-grams. *Expert Systems with Applications*, *124*, 156–163.

The shadowserver foundation. (2021). Retrieved 2021-02-08, from `https://www.shadowserver.org/`

Shi, Y., Chen, G., & Li, J. (2018). Malicious domain name detection based on extreme machine learning. *Neural Processing Letters*, *48*(3), 1347–1357.

Suarez-Tangil, G., Tapiador, J. E., Peris-Lopez, P., & Ribagorda, A. (2013). Evolution, detection and analysis of malware for smart devices. *IEEE Communications Surveys & Tutorials*, *16*(2), 961–987.

Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research. *How to Choose a Sampling Technique for Research (April 10, 2016)*.

Tanaka, Y., Akiyama, M., & Goto, A. (2017). Analysis of malware download sites by focusing on time series variation of malware. *Journal of computational science*, *22*, 301–313.

Thaichon, P., Lobo, A., Prentice, C., & Quach, T. N. (2014). The development of service quality dimensions for internet service providers: retaining customers of different usage patterns. *Journal of Retailing and Consumer Services*, *21*(6), 1047–1058.

Thaichon, P., & Quach, T. N. (2015). The relationship between service quality, satisfaction, trust, value, commitment and loyalty of internet service providers' customers. *Journal of Global Scholars of Marketing Science*, *25*(4), 295–313.

Thompson, N., McGill, T. J., & Wang, X. (2017). "security begins at home": Determinants of home computer and mobile device security behavior. *Computers & Security*, *70*, 376–391.

Tsai, H.-y. S., Jiang, M., Alhabash, S., LaRose, R., Rifon, N. J., & Cotten, S. R. (2016). Understanding online safety behaviors: A protection motivation theory perspective. *Computers & Security*, *59*, 138–150.

Vance, A., Siponen, M., & Pahnila, S. (2012). Motivating is security compliance: insights from habit and protection motivation theory. *Information & Management*, *49*(3-4), 190–198.

Van der Werff, L., Real, C., & Lynn, T. (2018). Individual trust and the internet. *DCU Online Research Access Service*.

Van Eeten, M., & Bauer, J. M. (2009). Emerging threats to internet security: Incentives, externalities and policy implications. *Journal of Contingencies and Crisis Management*, *17*(4), 221–232.

Van Eeten, M. J., & Bauer, J. M. (2008). Economics of malware: Security decisions, incentives and externalities. *OECD Science Technology and Industry Working Papers 2008/01*.

Van Schaik, P., Jeske, D., Onibokun, J., Coventry, L., Jansen, J., & Kusev, P. (2017). Risk perceptions of cyber-security and precautionary behaviour. *Computers in Human Behavior*, *75*, 547–559.

Venkatesh, V., Thong, J. Y., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the association for Information Systems*, *17*(5), 328–376.

Verstegen, S. (2019). Understanding the role of iot end users in miria-like bot remediation. *TU Delft Education Repository*.

Vixie, P., & Schryver, V. (2017). Response policy zones. *Internet Engenieering Task Force*, 10.

Vlajic, N., & Zhou, D. (2018). Iot as a land of opportunity for ddos hackers. *Computer*, *51*(7), 26–34.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the association for information systems*, *37*(1), 9.

Wang, K., Huang, C.-Y., Lin, S.-J., & Lin, Y.-D. (2011). A fuzzy pattern-based filtering algorithm for botnet detection. *Computer Networks*, *55*(15), 3275–3286.

Wang, P. A. (2010). Information security knowledge and behavior: An adapted model of technology acceptance. In *2010 2nd international conference on education technology and computer* (Vol. 2, pp. V2–364).

Wee, B. V., & Banister, D. (2016). How to write a literature review paper? *Transport Reviews*, *36*(2), 278–288.

West, R. (2008). The psychology of security. *Communications of the ACM*, *51*(4), 34–40.

What is a dns hijacking: Redirection attacks explained. (2019, Dec). Imperva. Retrieved from `https://www.imperva.com/learn/application-security/dns-hijacking-redirection/`

Woon, I., Tan, G.-W., & Low, R. (2005). A protection motivation theory approach to home wireless security..

Yaar, A., Perrig, A., & Song, D. (2003). Pi: A path identification mechanism to defend against ddos attacks. In *2003 symposium on security and privacy, 2003.* (pp. 93–107).

Yang, Y., Wu, L., Yin, G., Li, L., & Zhao, H. (2017). A survey on security and privacy issues in internet-of-things. *IEEE Internet of Things Journal*, *4*(5), 1250–1258.

Yoon, C., Hwang, J.-W., & Kim, R. (2012). Exploring factors that influence students' behaviors in information security. *Journal of information systems education*, *23*(4), 407–416.

Young, D. K., Carpenter, D., & McLeod, A. (2016). Malware avoidance motivations and behaviors: A technology threat avoidance replication. *AIS Transactions on Replication Research*, *2*(1), 8.

Yunus, M., Ibrahim, M., & Amir, F. (2018). The role of customer satisfaction and trust as mediation on the influence of service quality and corporate image to customer loyalty. *European Journal of Business and Management*, *10*(15), 121–128.

Zargar, S. T., Joshi, J., & Tipper, D. (2013). A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks. *IEEE communications surveys & tutorials*, *15*(4), 2046–2069.

Zaroo, P. (2002). A survey of ddos attacks and some ddos defense mechanisms. *Advanced Information Assurance (CS 626)*.

Zeidanloo, H. R., Shooshtari, M. J. Z., Amoli, P. V., Safari, M., & Zamani, M. (2010). A taxonomy of botnet detection techniques. In *2010 3rd international conference on computer science and information technology* (Vol. 2, pp. 158–162).

Zhou, Y., & Jiang, X. (2012). Dissecting android malware: Characterization and evolution. In *2012 ieee symposium on security and privacy* (pp. 95–109).

A

Literature Search

Article	Context	Method	Remarks
Feily 2009 survey	Botnets and botnet detection techniques	-	Characteristics (self-propagation, various C&C architectures) and life-cycle (infection, injection, connection, command and control, update and maintenance) of botnets. Taxonomy of botnet detection techniques; honeypots/honeynets and traffic analysis. Passive monitoring subdivided into signature-based, anomaly-based, DNS-based, and mining-based techniques. DNS-based techniques are most numerous and promising among the compared systems.
Zeidanloo 2010 taxonomy	Botnet detection techniques	-	Review of several detection techniques, primarily IRC and DNS-based. Taxonomy makes distinction between honeypots (attract threats to study) and intrusion detection systems (identify threats within a network). Anomaly-based and signature-based botnet detection (detection of unknown vs. well-known). Subdivided into host-based and network-based techniques, and further into active and passive monitoring. Host-based techniques tend to have significant false positives. Passive detection approaches appear to be the most accurate.
Alieyan 2017 survey	DNS based mitigation measures	-	Short section on the botnet lifecycle similar to that of Feily et. al. Taxonomy largely follows those set out by earlier works (honey DNS and IDS, subdivided into anomaly-based and signature-based techniques). Host-based and network based techniques, active and passive DNS techniques. Passive DNS techniques further subdivided into techniques based on the methods used to analyse the traffic (e.g. neural networks, decision trees, clustering). DNS blacklisting (similar to malwarefilter) and reputation-based systems prime examples of signature-based techniques.
Ramachandran 2006 revealing	Signature based techniques	DNS blackhole counter intelligence	Botmasters may attempt to assert whether their bots have been blacklisted or not; this may be done based on spatial or temporal relationships, and by third parties, a single host, or in a distributed manner. Bots were found to be conducting reconnaissance on IP addresses of bots in other botnets, providing opportunities for detectiong by examining DNS query graphs.
Antonakakis 2010 building	Signature based techniques	Dynamic reputation system	Employs passive DNS query data analysis to score the reputation of domains. Network-based features (number of IPs, geographical location, etc.), zone-based features (length of domain names, number of distinct top level domains, etc.) and evidence-based features (e.g. number of malware samples that contacted the domain). Evaluated in an ISP network based on traffic of 1.4 million users. High accuracy and low false positive rates. Strong reliance on historic data; not effective in qualifying new domains.
Otgonbold 2014 adapt	Active DNS techniques	Active probing of fast-flux domains	Employs data from domain zone files, mappings of IPs to ASNs, and public RDNS servers to (1) identify fast flux domains among DNS queries and (2) identify malicious domains among these fast flux domains. Despite some success in identifying (malicious) fast flux domains, relatively simple techniques can be employed to evade detection by the proposed system.
Ma 2015 accurate	Active DNS techniques	Active probing of DNS caches	Based on hyper-exponential distribution characteristics and time to live characteristics. Solution outperforms existing solutions based on an application of the solution to a large-scale, real-world DNS trace. System can be evaded, but only at a cost to the malicious activity. General evasion techniques tailored to the solution are still possible and, as the authors note, a challenge for any detection system.
Wang 2011 fuzzy	Passive DNS techniques	Fuzzy pattern recognition	Three step algorithm; traffic reduction, feature extraction, and fuzzy pattern recognition. Most common phenomena are noted to be failing DNS queries, similar query intervals, failed network flows, and similar payload sizes for different network flows. DNS features are used in conjunction with network traffic flow features in the fuzzy pattern recognition. Low-cost solution (computationally) in comparison to machine learning and statistical approaches, requiring only basic arithmetic rather than high-dimensional vectors. False-positive rate not insignificant.
Bilge 2011 exposure	Passive DNS techniques	Decision tree	15 features extracted from DNS traffic. Time-based features (lifespan, temporal similarity, access ratio, ...), answer-based features (dinstinct IPs, distinct countries, reverse query results, ...), TTL-based features (statistics over TTL value), and name-based features (percentage numerical characters, length of LMS). Evaluated on a real-world dataset of 100 billion DNS requests. High detection rate for malicious domains alongside a relatively high number of false positives.
Lee 2014 gmad	Passive DNS techniques	Graph algorithm	Use a graph construction and clustering techniques to establish relationships between (malicious) domains based on sequential correlation (e.g. domains being consistently queried at the same time or in a patterned order). The graph method makes it a scalable solution both temporally and spatially. Clusters are classified as malware or non-malware based on the occurrence of known malware domains among the cluster. Precision is reasonable, but perhaps the greatest benefit is an extremely low false positive rate (less than 0.30 per cent).
Shi 2018 malicious	Passive DNS techniques	Feedforward Neural Network	Employ a single-hidden-layer feed-forward neural network. Classification based on 9 features; domain name length, number of consecutive characters, entropy of domain, number of IP addresses and countries, TTL (avg. and std. dev.), doman life time and active time. Data set consisting of approximately 10 million DNS queries from a university (much smaller dataset than most other studies). Good accuracy and detection rate (comparable to or better than other ML-based efforts). Low training and testing times indicate possibility of real-time application.

Vixie 2017 response	Response policy zones	-	Decentralized, distributed nature of DNS complicates accountability and issue resolution. RPZ provides an open standard for DNSBL-like features in DNS. Rule-based system that allows the creation of fake or adjusted responses for purposes of, for example, creating walled gardens. Allows the elimination of malicious domains even if the responsible party does not suspend or terminate it. Notable downsides in the potential for censorship on the part of governments or organisations, politicalization of use, will become less effective over time.
Connery 2013 DNS	Response policy zones	-	RDNS configured to use RPZ uses zone files that contain (policy) information on DNS zones. Principally, these are locally defined zone files, which can nevertheless be obtained using zone file replication mechanisms (and thus can be defined by external parties). Common policies are NXDOMAIN, CNAME and PASSTHRU. The local zone data is queried before a full recursive query is made.

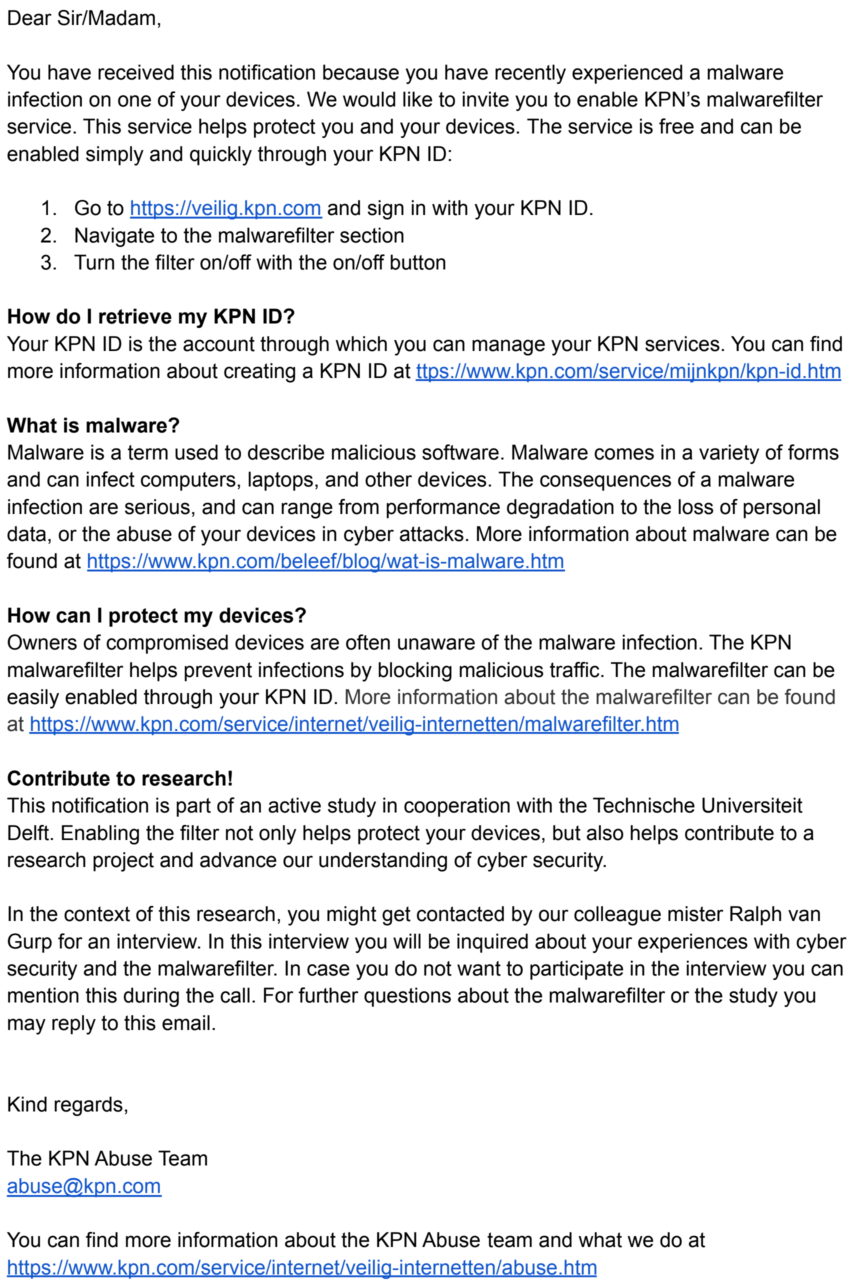
Article	Theory	Context	Significant	Insignificant	Remarks
Ng, B. Y., Kankanhalli, A., & Xu, Y. C. (2009). Studying users' computer security behavior: A health belief perspective.	HBM	Computer security behaviour	Susceptibility, benefits, self-efficacy	Barriers, cues to action, security orientation	Results indicate that subjects are not generally aware of the likelihood of a threats and therefore not capable of making educated decisions about the use of security measures. Understanding the benefit of security behaviour tends to be difficult.
Claar, C. L. (2011). The adoption of computer security: An analysis of home personal computer user behavior using the health belief model.	HBM	Protective technologies	Susceptibility, barriers, self-efficacy	Severity, benefits, cues to action	Contrary to the earlier work by Ng, specifically examines an end-user context. Non-probabilistic method of selection may have biased the findings. Further replication and expansion of the study required for definitive answers.
Dodel, M., & Mesch, G. (2017). Cyber-victimization preventive behavior: A health belief model approach.	HBM	Protective technologies	Susceptibility, severity, self-efficacy	Previous experiences	Significant disparity among demographic groups (gender, age, internet activity). Beliefs about the threat are found to be a major determinant in anti-virus use, alongside the need for solutions to be low-cost and low-disruptive with proven efficacy.
Kumar, N., Mohan, K., & Holowczak, R. (2008). Locking the door but leaving the computer vulnerable: Factors inhibiting home users' adoption of software firewalls.	TAM	Protective technologies	Perceived usefulness	Perceived ease of use	Examine a variety of constructs and relationships with regards to attitude and intention to adopt. Identify the `secondary' utility of a firewall relative to other applications of TAM. Study conducted among university students.
Wang, P. A. (2010, June). Information security knowledge and behavior: An adapted model of technology acceptance.	TAM	Protective technologies	Knowledge, Attitude, Intention to use	None	Subjects predominantly aged between 22 and 40, mostly students. Model does not contain any of the traditional TAM constructs, stead focussing on information, awareness, and experience.
Woon, I., Tan, G. W., & Low, R. (2005). A protection motivation theory approach to home wireless security.	PMT	Protective technologies (wireless security)	Perceived severity, response efficacy, response cost, self-efficacy	Perceived vulnerability	Survey of 189 home wireless network users. Significant results for coping appraisal factors but not for threat appraisal. Individuals with low self-confidence tend to also consider the response as less effective. Future research should include additional factors to improve the explanatory power.
Lee, Y., & Larsen, K. R. (2009). Threat or coping appraisal: determinants of SMB executives' decision to adopt anti-malware software.	PMT	Protective technologies (malware)	Perceived severity, perceived vulnerability, response efficacy, response cost, self-efficacy, social influence, vendor support, IT budget	Firm size	Intention is used as a mediator for adoption. All traditional PMT-constructs are found to be significant, most with p-values below 0.001, alongside a number of other factors (prominently; social influence).

Chenoweth, T., Minch, R., & Gattiker, T. (2009, January). Application of protection motivation theory to adoption of protective technologies.	PMT	Protective technologies (spyware)	Perceived vulnerability, perceived severity, response efficacy, response cost	Self-efficacy	Survey among undergraduate students. Insignificance of self-efficacy may be contextual; an assessment of the complexity of the response (response cost) rather than an individual's ability to perform the required actions.
Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: An empirical study.	PMT	Individual security behaviours	Perceived severity, self-efficacy, response efficacy, social influence	Perceived vulnerability	Subjects primarily aged between 18 and 29, facutly, staff, and students. Response efficacy and self-efficacy as mediators for threat vulnerability and severity. Response cost and actual behaviour not included in model.
Vance, A., Siponen, M., & Pahnila, S. (2012). Motivating IS security compliance: insights from habit and protection motivation theory..	PMT	IS Policy compliance	Perceived severity, maladaptive rewards, response efficacy, self-efficacy, response costs	Perceived vulnerability	Panel of 111 IS security experts from a single organisation, final dataset contains 54 responses. Study centered around the impact of habit. Similar to other studies the significant of certain hypotheses may be impacted by the organisational context.
Ifinedo, P. (2012). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory.	PMT	IS Policy compliance	Perceived vulnerability, response efficacy, self-efficacy, attitude towards compliance, subjective norms	Response cost, perceived severity	Survey among 124 business managers and IS professionals. Fusion of PMT and TPB-elements such as subjective norm. Non-significant relations may be related by the organizational context of the study. Qualitative methods may be used to provide greater insights.
Crossler, R. E., Long, J. H., Loraas, T. M., & Trinkle, B. S. (2014). Understanding compliance with bring your own device policies utilizing protection motivation theory: Bridging the intention-behavior gap.	PMT	Individual security behaviours	Perceived severity, perceived vulnerability, response efficacy, self-efficacy	Reponse cost	Survey among 81 graduate students in a non-IT related discipline. Suggest a ground theory or comparable approach to understand what motivates users to perform security behaviours, as awell as expand the model to other contexts and threats.
Hanus, B., & Wu, Y. A. (2016). Impact of users' security awareness on desktop security behavior: A protection motivation theory perspective.	PMT	Desktop security behaviours	Self-efficacy, response efficacy, awareness	Perceived severity, perceived vulnerability, response costs	Survey among undergraduate students of multiple disciplines. Awareness considered to be a critical factor in assessing both threats and protection mechanisms. Threat appraisal is not a significant predictor of behaviour. Assess only financial aspects of response cost.
Tsai, H. Y. S., Jiang, M., Alhabash, S., LaRose, R., Rifon, N. J., & Cotten, S. R. (2016). Understanding online safety behaviors: A protection motivation theory perspective.	PMT	Security behaviour (online)	Prior experience, response efficacy, subjective norm, response cost, safety habits, responsibility	Perceived severity, perceived vulnerability, self-efficacy, perceived security support	Survey among Amazon Mechanical Turk users, skewed towards a young, highly-educated population. The incorporation of factors such as prior experience and subjective norms added significant explanatory power. Understanding actual behaviours in addition to intentions may yield additional insights.

Ophoff, J., & Lakay, M. (2018, August). Mitigating the ransomware threat: a protection motivation theory approach.	PMT	Protective technologies (ransomware)	Fear, maladaptive rewards, self-efficacy, response cost	Perceived severity, perceived vulnerability, response efficacy	Perceived severity and vulnerability are indirectly significant (mediated by fear). Limited and homogenous sample of university students and professors. Protection motivation as a combination of both technology (anti-malware software) and behaviours (backing up data).
Martens, M., De Wolf, R., & De Marez, L. (2019). Investigating and comparing the predictors of the intention towards taking security measures against malware, scams and cybercrime in general.	PMT	Protective technologies (scams, malware)	Response efficacy, perceived vulnerability, perceived severity, subjective norm	Self-efficacy	Differences in significance between models for various threat types (scams, malware, and general cyber threats). Threat awareness is found to be an important factor in perceived severity, possibly because of the technical complexity associated with malware threats. Conversely, knowledge and awareness may result in individuals overestimating their abilities and understimating the likelihood of being victimized.
Liang, H., & Xue, Y. L. (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective.	TTAT	Protective technologies (spyware)	Perceived severity, perceived susceptibility, safeguard effectiveness, safeguard cost, self-efficacy	None	There is a necessity for subjects to be be aware of the likelihood a threat; to understand that they exist and are avoidable. Conventional approaches towards security rely on end-users as passive subjects rather than active participants. Indication of negative interationcs between threat perception and safeguard efficacy.
Arachchilage, N. A. G., & Love, S. (2014). Security awareness of computer users: A phishing threat avoidance perspective.	TTAT	Security behaviour (phishing)	Self-efficacy (procedural and conceptual knowledge)	None (limited model)	Study among 18-25 year olds, restricted to procedural and conceptual knowledge and their effect on self-efficacy, avoidance motivation, and avoidance behaviour. Both were found to significant affect self-efficacy and self-efficacy, in turn, is a significant predictor of avoidance motivation.
Young, D. K., Carpenter, D., & McLeod, A. (2016). Malware avoidance motivations and behaviors: A technology threat avoidance replication.	TTAT	Protective technologies (malware)	Perceived severity, safeguard effectiveness, safeguard cost, self-efficacy	Perceived susceptibility	Survey among university students may have biased the findings. Findings suggest that some significant predictors for avoidance motivation are missing from the model, predominantly the inclusion of risk and social factors.

B

Malwarefilter Notification



Dear Sir/Madam,

You have received this notification because you have recently experienced a malware infection on one of your devices. We would like to invite you to enable KPN's malwarefilter service. This service helps protect you and your devices. The service is free and can be enabled simply and quickly through your KPN ID:

1. Go to https://veilig.kpn.com and sign in with your KPN ID.
2. Navigate to the malwarefilter section
3. Turn the filter on/off with the on/off button

How do I retrieve my KPN ID?
Your KPN ID is the account through which you can manage your KPN services. You can find more information about creating a KPN ID at ttps://www.kpn.com/service/mijnkpn/kpn-id.htm

What is malware?
Malware is a term used to describe malicious software. Malware comes in a variety of forms and can infect computers, laptops, and other devices. The consequences of a malware infection are serious, and can range from performance degradation to the loss of personal data, or the abuse of your devices in cyber attacks. More information about malware can be found at https://www.kpn.com/beleef/blog/wat-is-malware.htm

How can I protect my devices?
Owners of compromised devices are often unaware of the malware infection. The KPN malwarefilter helps prevent infections by blocking malicious traffic. The malwarefilter can be easily enabled through your KPN ID. More information about the malwarefilter can be found at https://www.kpn.com/service/internet/veilig-internetten/malwarefilter.htm

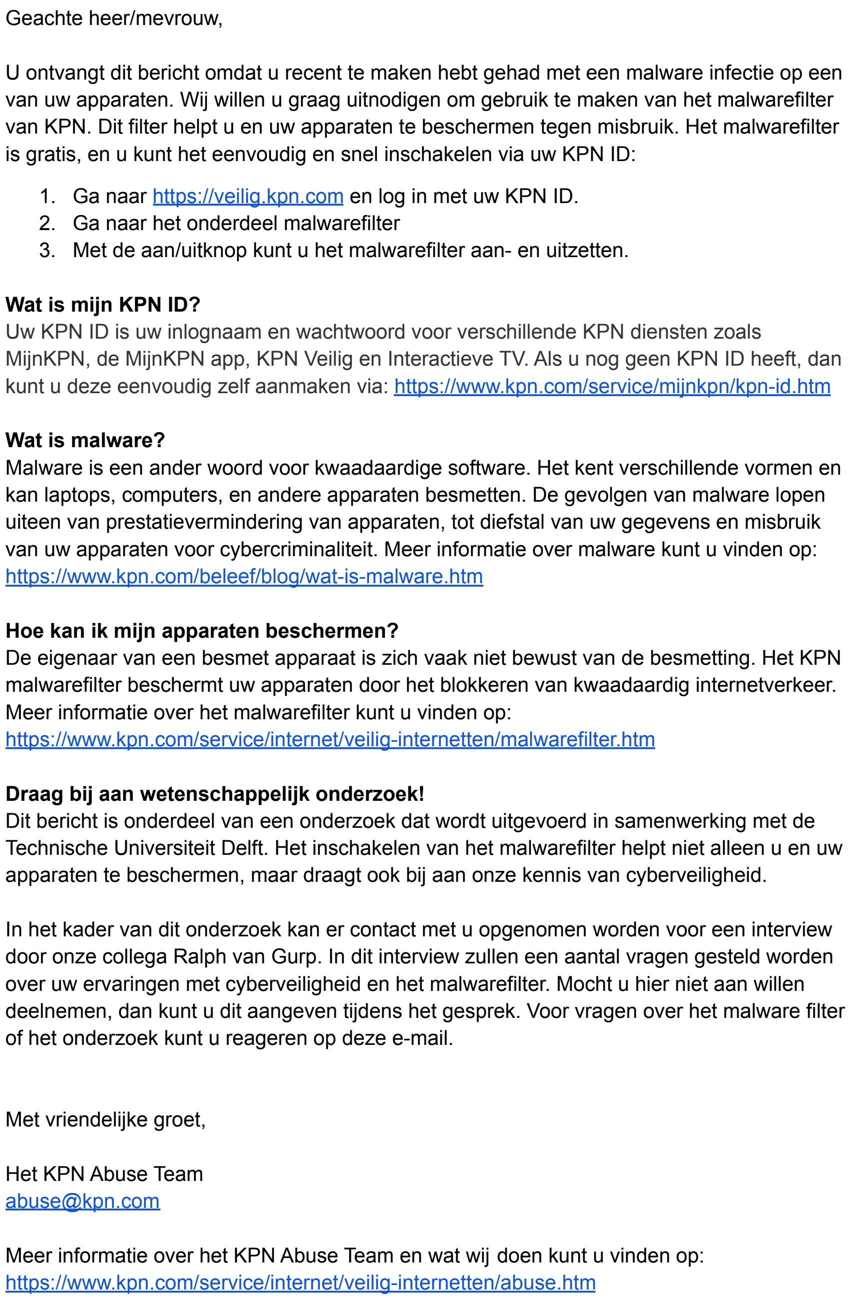
Contribute to research!
This notification is part of an active study in cooperation with the Technische Universiteit Delft. Enabling the filter not only helps protect your devices, but also helps contribute to a research project and advance our understanding of cyber security.

In the context of this research, you might get contacted by our colleague mister Ralph van Gurp for an interview. In this interview you will be inquired about your experiences with cyber security and the malwarefilter. In case you do not want to participate in the interview you can mention this during the call. For further questions about the malwarefilter or the study you may reply to this email.

Kind regards,

The KPN Abuse Team
abuse@kpn.com

You can find more information about the KPN Abuse team and what we do at
https://www.kpn.com/service/internet/veilig-internetten/abuse.htm



Geachte heer/mevrouw,

U ontvangt dit bericht omdat u recent te maken hebt gehad met een malware infectie op een van uw apparaten. Wij willen u graag uitnodigen om gebruik te maken van het malwarefilter van KPN. Dit filter helpt u en uw apparaten te beschermen tegen misbruik. Het malwarefilter is gratis, en u kunt het eenvoudig en snel inschakelen via uw KPN ID:

1. Ga naar https://veilig.kpn.com en log in met uw KPN ID.
2. Ga naar het onderdeel malwarefilter
3. Met de aan/uitknop kunt u het malwarefilter aan- en uitzetten.

Wat is mijn KPN ID?
Uw KPN ID is uw inlognaam en wachtwoord voor verschillende KPN diensten zoals MijnKPN, de MijnKPN app, KPN Veilig en Interactieve TV. Als u nog geen KPN ID heeft, dan kunt u deze eenvoudig zelf aanmaken via: https://www.kpn.com/service/mijnkpn/kpn-id.htm

Wat is malware?
Malware is een ander woord voor kwaadaardige software. Het kent verschillende vormen en kan laptops, computers, en andere apparaten besmetten. De gevolgen van malware lopen uiteen van prestatievermindering van apparaten, tot diefstal van uw gegevens en misbruik van uw apparaten voor cybercriminaliteit. Meer informatie over malware kunt u vinden op: https://www.kpn.com/beleef/blog/wat-is-malware.htm

Hoe kan ik mijn apparaten beschermen?
De eigenaar van een besmet apparaat is zich vaak niet bewust van de besmetting. Het KPN malwarefilter beschermt uw apparaten door het blokkeren van kwaadaardig internetverkeer. Meer informatie over het malwarefilter kunt u vinden op: https://www.kpn.com/service/internet/veilig-internetten/malwarefilter.htm

Draag bij aan wetenschappelijk onderzoek!
Dit bericht is onderdeel van een onderzoek dat wordt uitgevoerd in samenwerking met de Technische Universiteit Delft. Het inschakelen van het malwarefilter helpt niet alleen u en uw apparaten te beschermen, maar draagt ook bij aan onze kennis van cyberveiligheid.

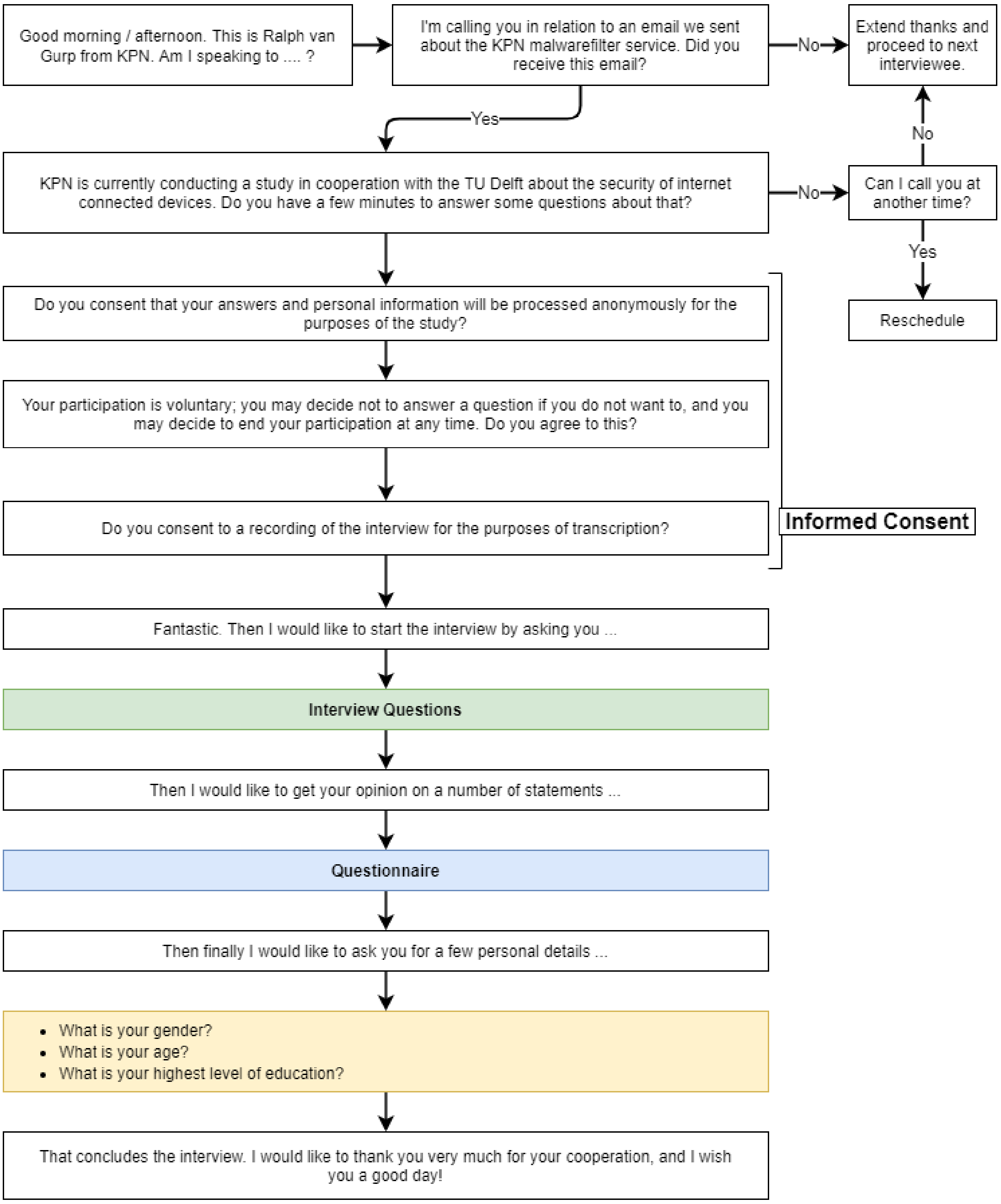
In het kader van dit onderzoek kan er contact met u opgenomen worden voor een interview door onze collega Ralph van Gurp. In dit interview zullen een aantal vragen gesteld worden over uw ervaringen met cyberveiligheid en het malwarefilter. Mocht u hier niet aan willen deelnemen, dan kunt u dit aangeven tijdens het gesprek. Voor vragen over het malware filter of het onderzoek kunt u reageren op deze e-mail.

Met vriendelijke groet,

Het KPN Abuse Team
abuse@kpn.com

Meer informatie over het KPN Abuse Team en wat wij doen kunt u vinden op:
https://www.kpn.com/service/internet/veilig-internetten/abuse.htm

C
Pilot interview protocol



Good morning / afternoon. This is Ralph van Gurp from KPN. Am I speaking to .... ?
I'm calling you in relation to an email we sent about the KPN malwarefilter service. Did you receive this email?
No
Extend thanks and proceed to next interviewee.
Yes
KPN is currently conducting a study in cooperation with the TU Delft about the security of internet connected devices. Do you have a few minutes to answer some questions about that?
No
Can I call you at another time?
No
Yes
Reschedule
Do you consent that your answers and personal information will be processed anonymously for the purposes of the study?
Your participation is voluntary; you may decide not to answer a question if you do not want to, and you may decide to end your participation at any time. Do you agree to this?
Do you consent to a recording of the interview for the purposes of transcription?
Informed Consent
Fantastic. Then I would like to start the interview by asking you ...
Interview Questions
Then I would like to get your opinion on a number of statements ...
Questionnaire
Then finally I would like to ask you for a few personal details ...
What is your gender?
What is your age?
What is your highest level of education?
That concludes the interview. I would like to thank you very much for your cooperation, and I wish you a good day!

Figure C.1: Overview of the interview protocol (pre-pilot).

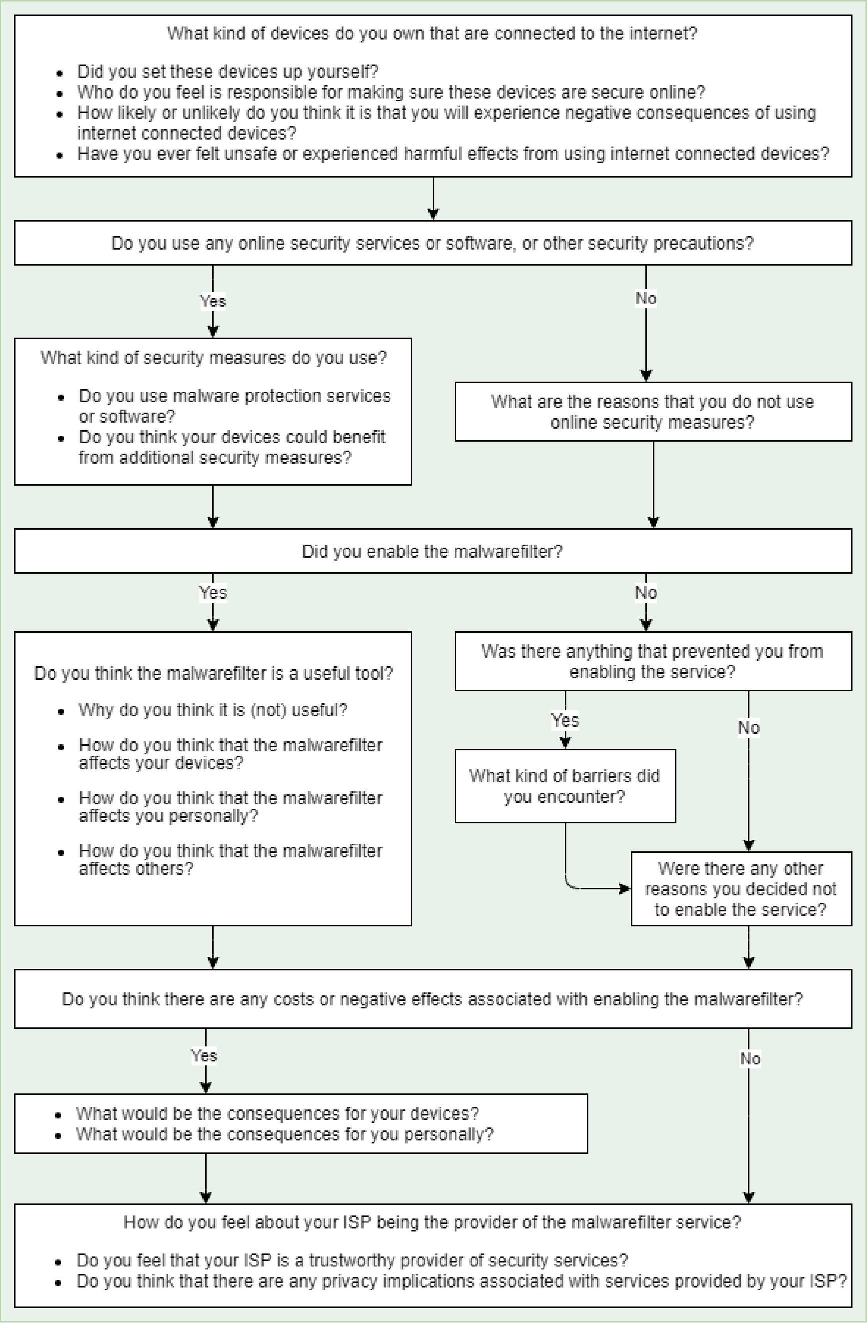
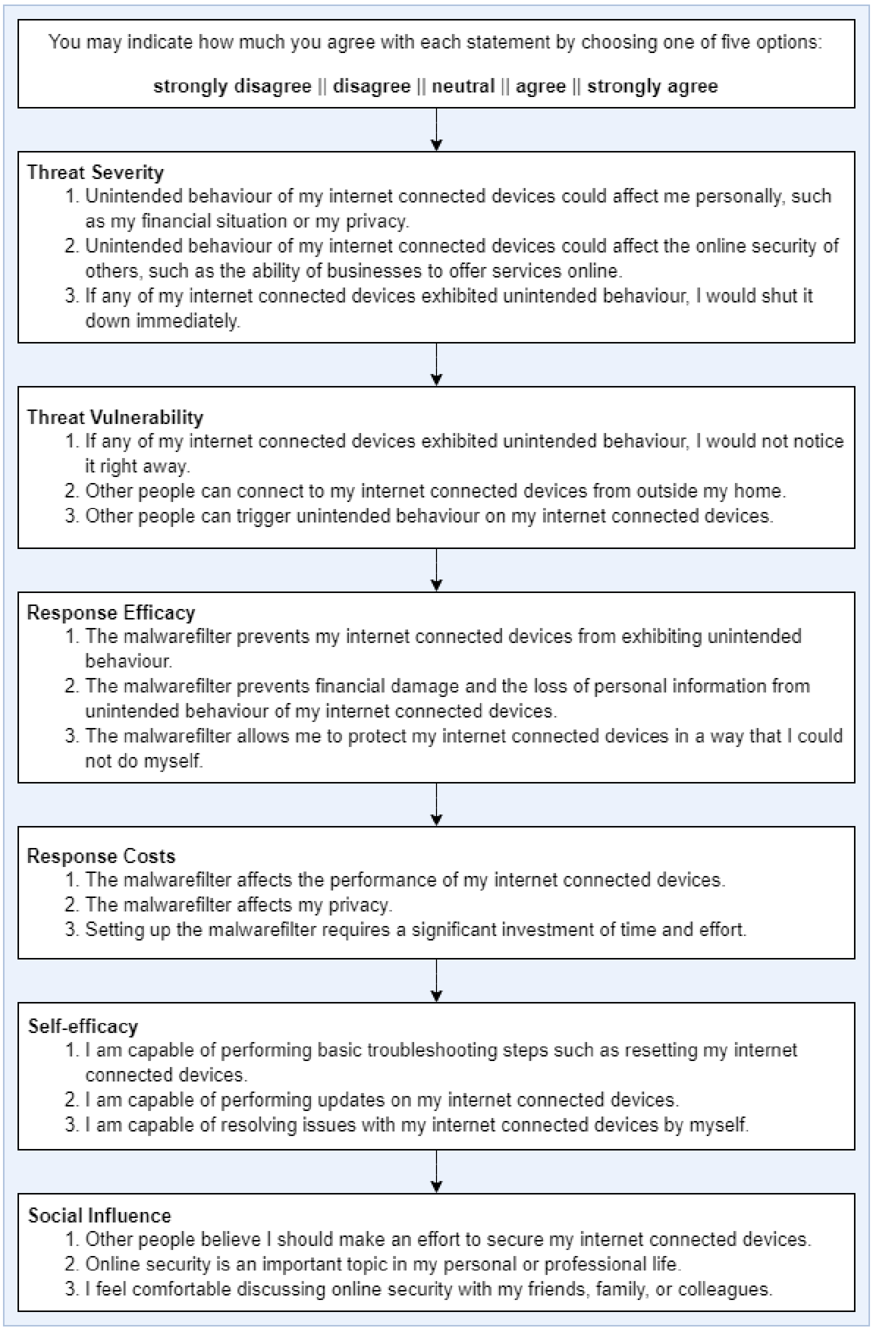

What kind of devices do you own that are connected to the internet?
Did you set these devices up yourself?
Who do you feel is responsible for making sure these devices are secure online?
How likely or unlikely do you think it is that you will experience negative consequences of using internet connected devices?
Have you ever felt unsafe or experienced harmful effects from using internet connected devices?
Do you use any online security services or software, or other security precautions?
Yes
No
What kind of security measures do you use?
Do you use malware protection services or software?
Do you think your devices could benefit from additional security measures?
What are the reasons that you do not use online security measures?
Did you enable the malwarefilter?
Yes
No
Do you think the malwarefilter is a useful tool?
Why do you think it is (not) useful?
How do you think that the malwarefilter affects your devices?
How do you think that the malwarefilter affects you personally?
How do you think that the malwarefilter affects others?
Was there anything that prevented you from enabling the service?
Yes
No
What kind of barriers did you encounter?
Were there any other reasons you decided not to enable the service?
Do you think there are any costs or negative effects associated with enabling the malwarefilter?
Yes
No
What would be the consequences for your devices?
What would be the consequences for you personally?
How do you feel about your ISP being the provider of the malwarefilter service?
Do you feel that your ISP is a trustworthy provider of security services?
Do you think that there are any privacy implications associated with services provided by your ISP?

Figure C.2: Question section of the interview protocol (pre-pilot).



You may indicate how much you agree with each statement by choosing one of five options:

strongly disagree || **disagree** || **neutral** || **agree** || **strongly agree**

Threat Severity

1. Unintended behaviour of my internet connected devices could affect me personally, such as my financial situation or my privacy.
2. Unintended behaviour of my internet connected devices could affect the online security of others, such as the ability of businesses to offer services online.
3. If any of my internet connected devices exhibited unintended behaviour, I would shut it down immediately.

Threat Vulnerability

1. If any of my internet connected devices exhibited unintended behaviour, I would not notice it right away.
2. Other people can connect to my internet connected devices from outside my home.
3. Other people can trigger unintended behaviour on my internet connected devices.

Response Efficacy

1. The malwarefilter prevents my internet connected devices from exhibiting unintended behaviour.
2. The malwarefilter prevents financial damage and the loss of personal information from unintended behaviour of my internet connected devices.
3. The malwarefilter allows me to protect my internet connected devices in a way that I could not do myself.

Response Costs

1. The malwarefilter affects the performance of my internet connected devices.
2. The malwarefilter affects my privacy.
3. Setting up the malwarefilter requires a significant investment of time and effort.

Self-efficacy

1. I am capable of performing basic troubleshooting steps such as resetting my internet connected devices.
2. I am capable of performing updates on my internet connected devices.
3. I am capable of resolving issues with my internet connected devices by myself.

Social Influence

1. Other people believe I should make an effort to secure my internet connected devices.
2. Online security is an important topic in my personal or professional life.
3. I feel comfortable discussing online security with my friends, family, or colleagues.

Figure C.3: Statement section of the interview protocol (pre-pilot).

D

Revised interview protocol

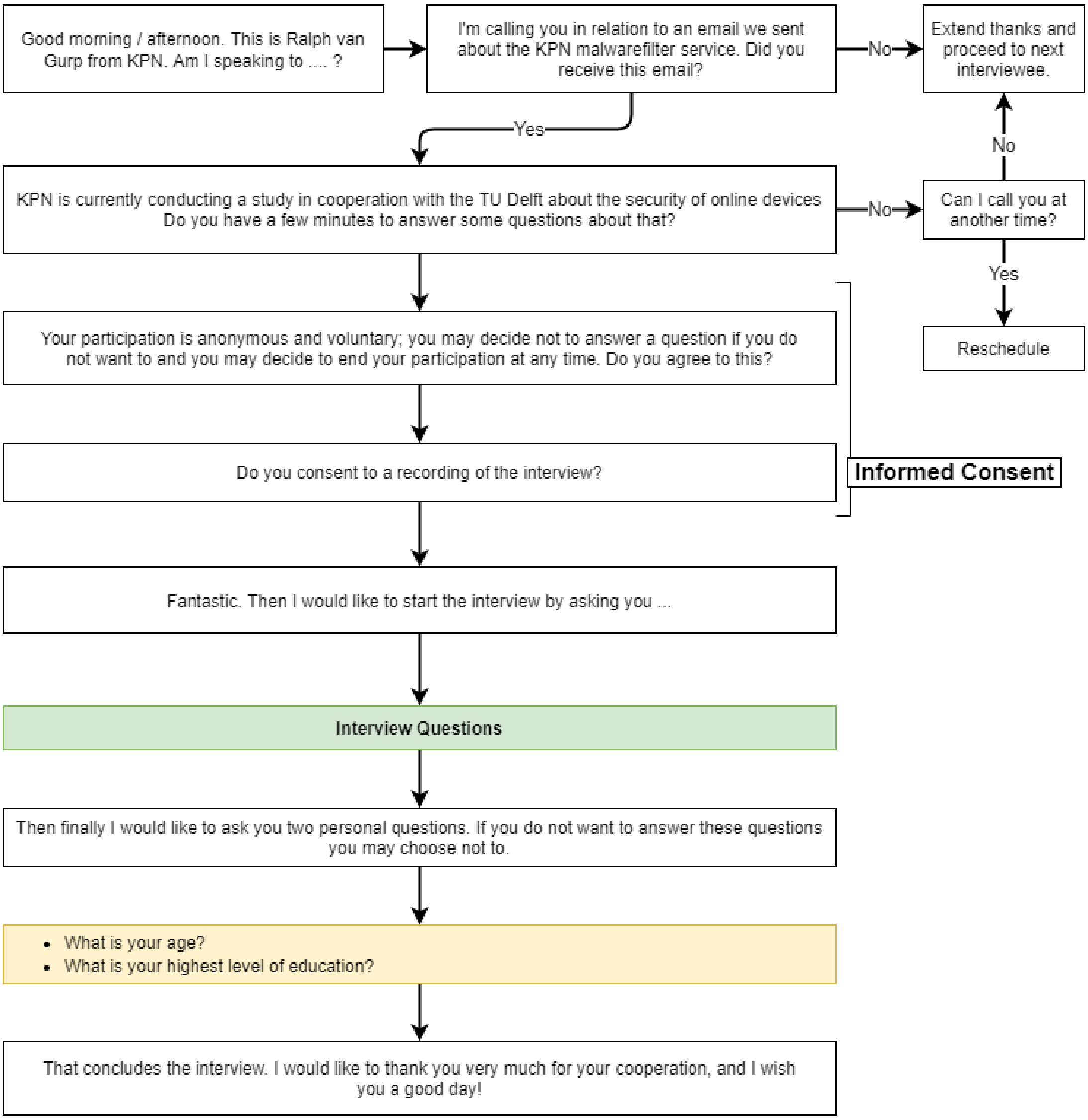

Good morning / afternoon. This is Ralph van Gurp from KPN. Am I speaking to .... ?
I'm calling you in relation to an email we sent about the KPN malwarefilter service. Did you receive this email?
No
Extend thanks and proceed to next interviewee.
Yes
No
KPN is currently conducting a study in cooperation with the TU Delft about the security of online devices Do you have a few minutes to answer some questions about that?
No
Can I call you at another time?
Yes
Reschedule
Your participation is anonymous and voluntary; you may decide not to answer a question if you do not want to and you may decide to end your participation at any time. Do you agree to this?
Do you consent to a recording of the interview?
Informed Consent
Fantastic. Then I would like to start the interview by asking you ...
Interview Questions
Then finally I would like to ask you two personal questions. If you do not want to answer these questions you may choose not to.
What is your age?
What is your highest level of education?
That concludes the interview. I would like to thank you very much for your cooperation, and I wish you a good day!

Figure D.1: Revised interview protocol (post-pilot).

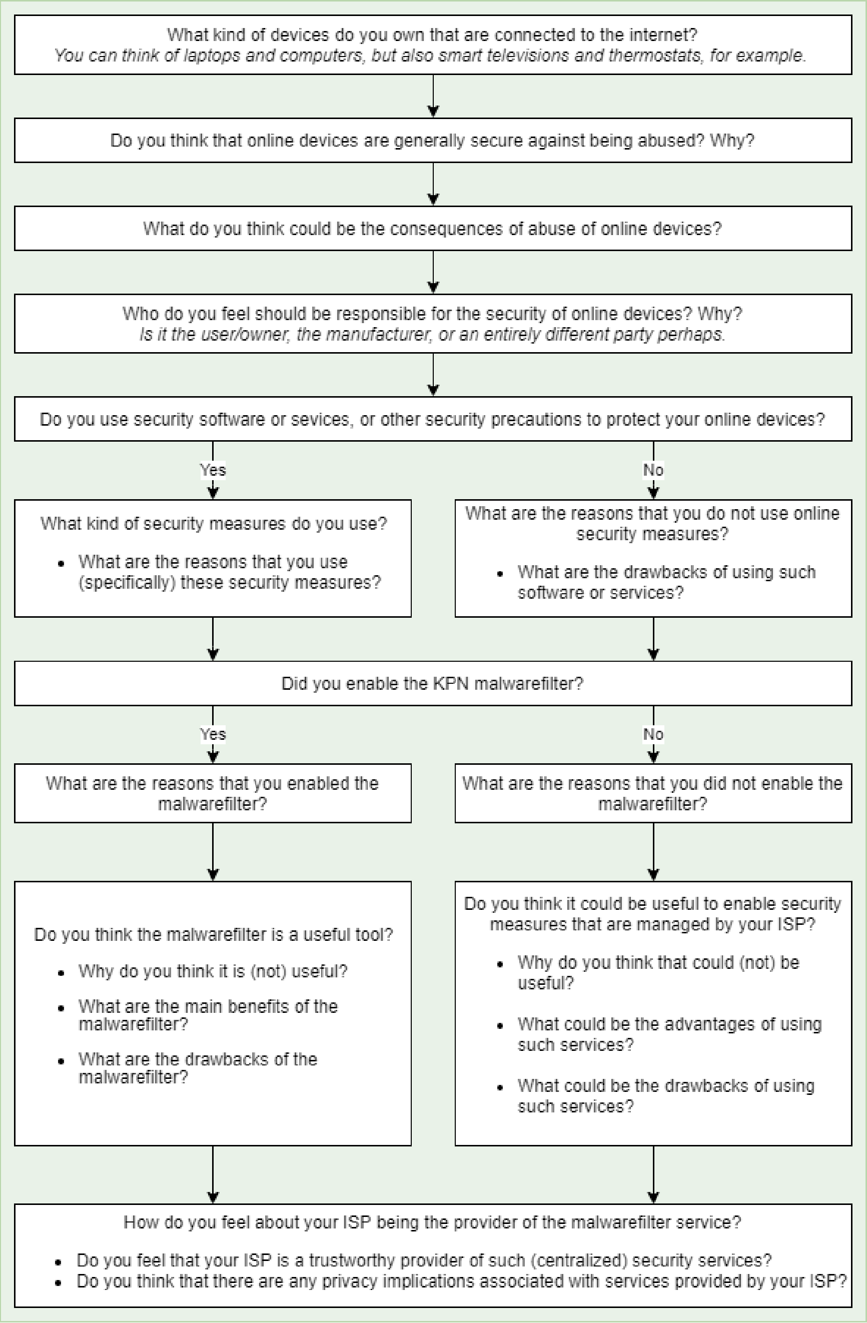

What kind of devices do you own that are connected to the internet?
You can think of laptops and computers, but also smart televisions and thermostats, for example.
Do you think that online devices are generally secure against being abused? Why?
What do you think could be the consequences of abuse of online devices?
Who do you feel should be responsible for the security of online devices? Why?
Is it the user/owner, the manufacturer, or an entirely different party perhaps.
Do you use security software or sevices, or other security precautions to protect your online devices?
Yes
No
What kind of security measures do you use?
What are the reasons that you use (specifically) these security measures?
What are the reasons that you do not use online security measures?
What are the drawbacks of using such software or services?
Did you enable the KPN malwarefilter?
Yes
No
What are the reasons that you enabled the malwarefilter?
What are the reasons that you did not enable the malwarefilter?
Do you think the malwarefilter is a useful tool?
Why do you think it is (not) useful?
What are the main benefits of the malwarefilter?
What are the drawbacks of the malwarefilter?
Do you think it could be useful to enable security measures that are managed by your ISP?
Why do you think that could (not) be useful?
What could be the advantages of using such services?
What could be the drawbacks of using such services?
How do you feel about your ISP being the provider of the malwarefilter service?
Do you feel that your ISP is a trustworthy provider of such (centralized) security services?
Do you think that there are any privacy implications associated with services provided by your ISP?

Figure D.2: Revised question section of the interview protocol (post-pilot).

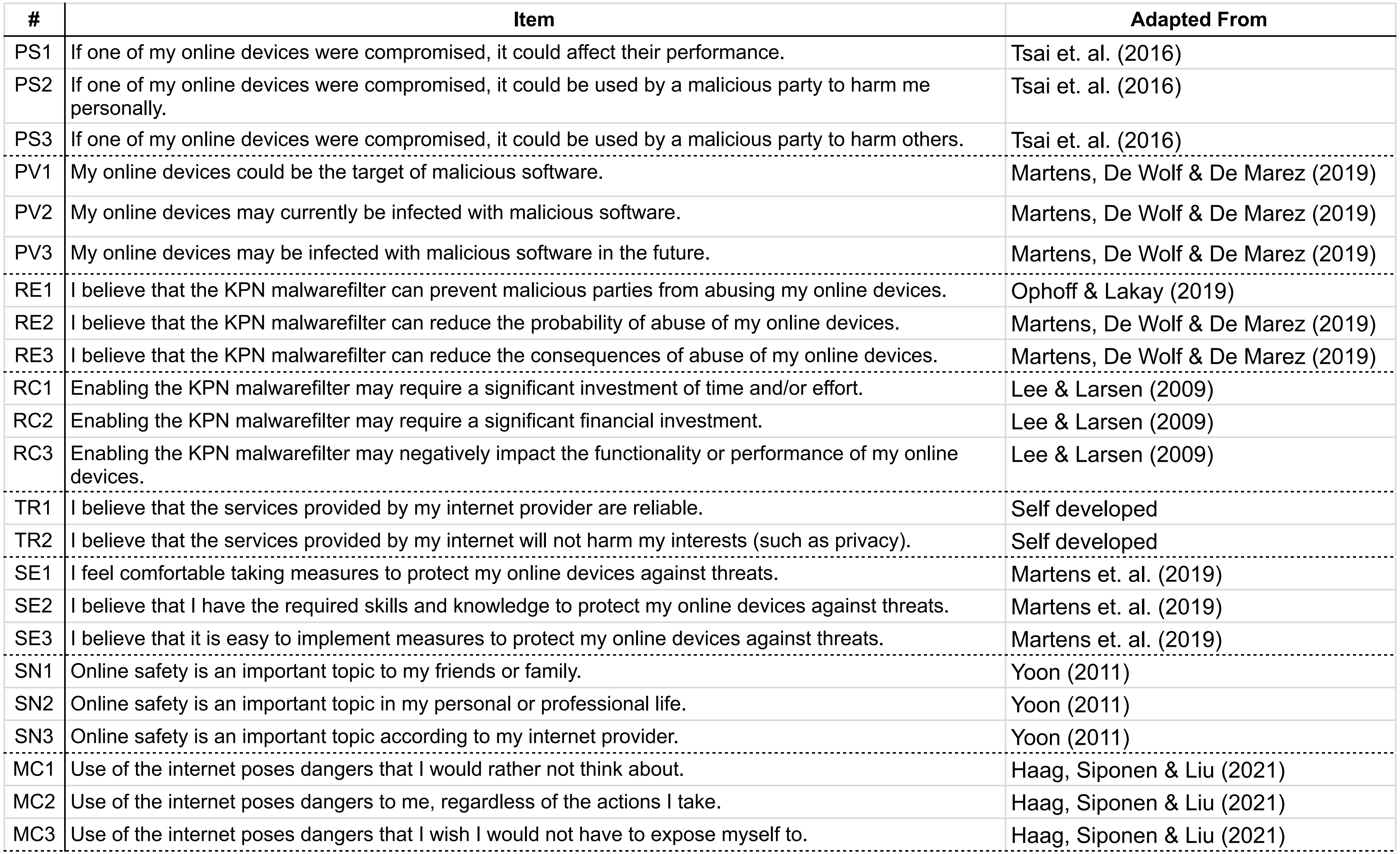
E

Pilot questionnaire

#	Item	Cronbach's Alpha	Internal Consistency	Remarks
PS1	If one of my online devices were compromised, it could affect their performance.	0,652	Acceptable	Excluding PS2 yields an alpha of 0,706
PS2	If one of my online devices were compromised, it could be used by a malicious party to harm me personally.			
PS3	If one of my online devices were compromised, it could be used by a malicious party to harm others.			
PV1	My online devices are vulnerable to being compromised by a malicious party.	0,000	Unacceptable	Revisited the survey items to be more strongly in line with earlier works
PV2	My online devices are likely to be compromised by a malicious party.			
RE1	I believe that the KPN malwarefilter can prevent malicious parties from abusing my online devices.	0,700	Acceptable	Excluding RE1 yields an alpha of 0,933
RE2	I believe that the KPN malwarefilter can reduce the probability of abuse of my online devices.			
RE3	I believe that the KPN malwarefilter can reduce the consequences of abuse of my online devices.			
RC1	Enabling the KPN malwarefilter may require a significant investment of time and/or effort.	0,556	Acceptable	Excluding RC3 yields an alpha of 1,000
RC2	Enabling the KPN malwarefilter may require a significant financial investment.			
RC3	Enabling the KPN malwarefilter may negatively impact the functionality or performance of my online devices.			
TR1	I believe that the services provided by my internet provider are reliable.	1,000	Excellent	None
TR2	I believe that the services provided by my internet will not harm my interests (such as privacy).			
SE1	I believe that I can make informed decisions about the use of security measures.	0,656	Acceptable	Excluding SE1 yields an alpha of 0,933. Revisited SE1.
SE2	I believe that I have sufficient knowledge to protect my online devices against threats.			
SE3	I believe that it is easy to implement measures to protect my online devices against threats.			
SN1	Online safety is an important topic to my friends or family.	0,625	Acceptable	Excluding SN2 yields an alpha of 0,833
SN2	Online safety is an important topic in my personal or professional life.			
SN3	Online safety is an important topic according to my internet provider.			
MC1	Use of the internet poses dangers that I would rather not think about.	0,844	Good	Added an additional item/dimension
MC2	Use of the internet poses dangers to me, regardless of the actions I take.			

F

Revised questionnaire



#	Item	Adapted From
PS1	If one of my online devices were compromised, it could affect their performance.	Tsai et. al. (2016)
PS2	If one of my online devices were compromised, it could be used by a malicious party to harm me personally.	Tsai et. al. (2016)
PS3	If one of my online devices were compromised, it could be used by a malicious party to harm others.	Tsai et. al. (2016)
PV1	My online devices could be the target of malicious software.	Martens, De Wolf & De Marez (2019)
PV2	My online devices may currently be infected with malicious software.	Martens, De Wolf & De Marez (2019)
PV3	My online devices may be infected with malicious software in the future.	Martens, De Wolf & De Marez (2019)
RE1	I believe that the KPN malwarefilter can prevent malicious parties from abusing my online devices.	Ophoff & Lakay (2019)
RE2	I believe that the KPN malwarefilter can reduce the probability of abuse of my online devices.	Martens, De Wolf & De Marez (2019)
RE3	I believe that the KPN malwarefilter can reduce the consequences of abuse of my online devices.	Martens, De Wolf & De Marez (2019)
RC1	Enabling the KPN malwarefilter may require a significant investment of time and/or effort.	Lee & Larsen (2009)
RC2	Enabling the KPN malwarefilter may require a significant financial investment.	Lee & Larsen (2009)
RC3	Enabling the KPN malwarefilter may negatively impact the functionality or performance of my online devices.	Lee & Larsen (2009)
TR1	I believe that the services provided by my internet provider are reliable.	Self developed
TR2	I believe that the services provided by my internet will not harm my interests (such as privacy).	Self developed
SE1	I feel comfortable taking measures to protect my online devices against threats.	Martens et. al. (2019)
SE2	I believe that I have the required skills and knowledge to protect my online devices against threats.	Martens et. al. (2019)
SE3	I believe that it is easy to implement measures to protect my online devices against threats.	Martens et. al. (2019)
SN1	Online safety is an important topic to my friends or family.	Yoon (2011)
SN2	Online safety is an important topic in my personal or professional life.	Yoon (2011)
SN3	Online safety is an important topic according to my internet provider.	Yoon (2011)
MC1	Use of the internet poses dangers that I would rather not think about.	Haag, Siponen & Liu (2021)
MC2	Use of the internet poses dangers to me, regardless of the actions I take.	Haag, Siponen & Liu (2021)
MC3	Use of the internet poses dangers that I wish I would not have to expose myself to.	Haag, Siponen & Liu (2021)

#	Question	Answers
1	How often do you actively use the following devices? Computers (e.g. desktop computer, game console) Mobile devices (e.g. laptop, tablet, smartphone) Smart-home devices (e.g. smart speaker, intelligent thermostat, smart lighting)	Do not own / Daily / Weekly / Monthly / Less often than monthly
2	To what extend are you worried about possible dangers to your online devices such as malicious software or data theft?	Not concerned / Somewhat concerned / Highly concerned
3	Which of the following security incidents have you ever had an experience with? Data theft (compromised authentication details or other sensitive information) Phishing (deceptive messages intended to persuade you to perform certain actions) Malware (Malicious software installed on one of your devices)	Multiple Choice
4	How often do you perform the following security actions? Installing security updates Changing passwords Backing up important data	Never / Seldom / Sometimes / Regularly / Often
5	What kind of security software or services do you use to protect your online devices? Device-default security measures (e.g. Microsoft Defender) Services provided by your ISP (e.g. KPN veilig) Services provided by third parties (e.g. Norton, McAfee, MalwareBytes) None of the above	Multiple Choice
6	Are you familiar with the KPN malwarefilter? Do you currently use the KPN malwarefilter?	Yes / No Yes / No / I don’t know
7	How likely is it that you will enable or keep using the KPN malwarefilter in the future?	Highly unlikely / Unlikely / Likely / Highly likely
8	What is your age (in years)?	Numeric
9	What is your gender?	Male / Female / Other / Prefer not to say
10	What is your highest level of education?	Secondary education / Vocational education / Higher education
11	Did you receive an education in the field of IT, or are you or have you previously been employed in the IT sector?	Yes / No
12	How experienced do you consider yourself in the use of technology?	Inexperienced / Experienced / Highly experienced

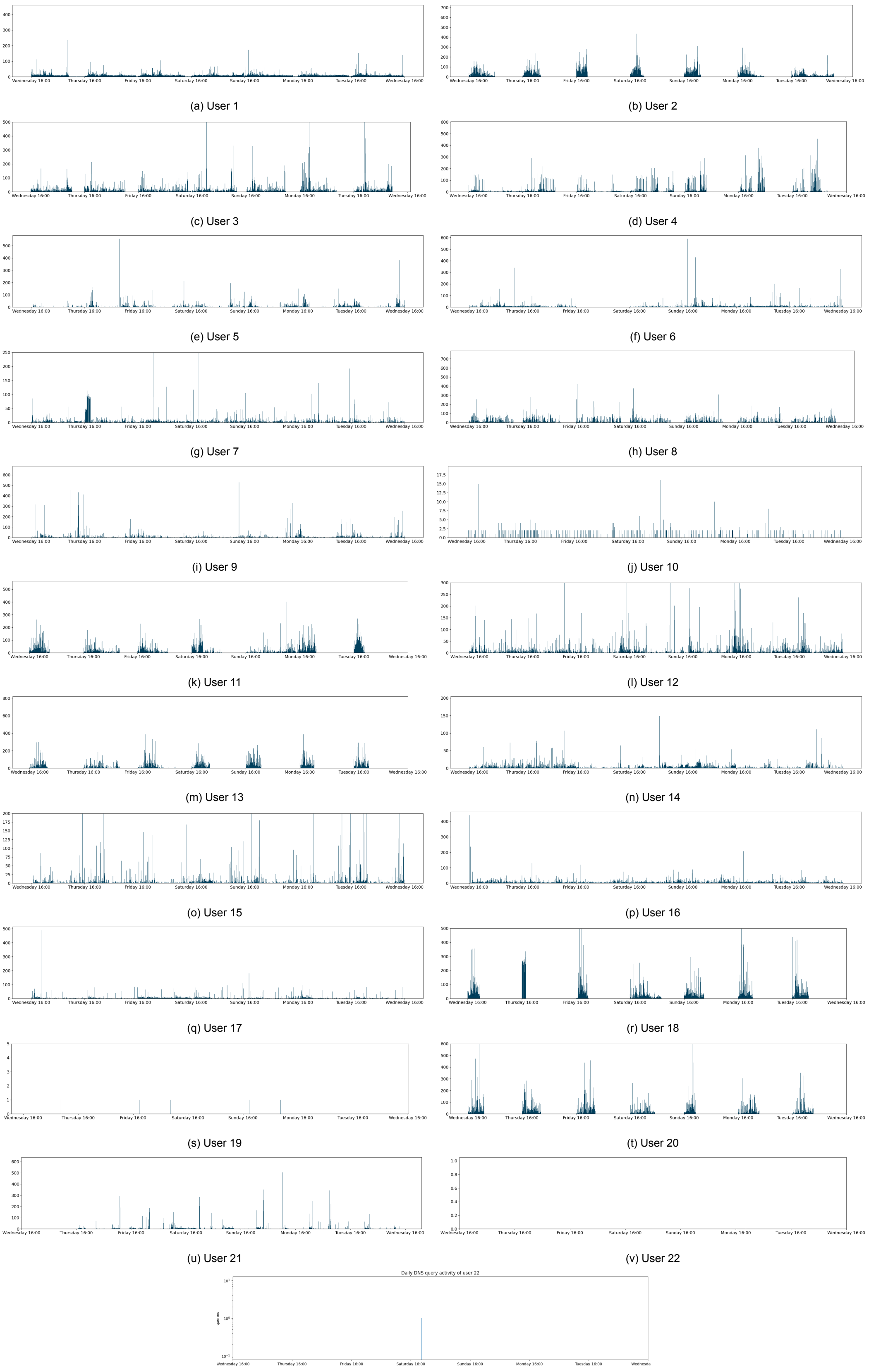
G

Codebook

Topic	Category	Code	Description
Assessing online threats	*Vulnerability to threats*	familiarity with security incidents	The degree to which an individual has first or second-hand experienced security incidents, or is aware of such incidents due to media coverage.
		difficult to assess threats	Expressing difficulty in assessing the dangers posed by online threats, for example because of a lack of knowledge about the threatscape.
		cannot rule out threats entirely	A perceived inability to prevent or notably mitigate the dangers of online threats.
		not a likely target of threats	The view that the individual is unlikely to be a (deliberate) target of malicious actors or activity.
	Consequences of threats	consequences for user	Identifying the possible consequences of online threats for the user or owner of a device, such as data loss or identity theft.
		consequences for devices	Identifying the possible consequences of online threats for a compromised device, affecting its performance or functioning.
		consequences for others	Identifying the possible consequences of online threats for others, negatively affecting parties beyond the user or owner of a compromised device.
Assessing device security	*Assessing security*	insufficiently capable of assessing	Expressing difficulty in assessing the security of internet connected devices, for example because of a lack of knowledge about cyber security.
		depends on device characteristics	The degree to which the perceived security of a device depends on characteristics such as its operating system or age.
		doubts about device security (undefined)	Expressing doubts about internet connected devices being adequately secure out-of-the-box or even after employing security measures.
	Necessity of security	distrust of manufacturer default configuration	Distrust of the security of the default configuration or security measures afforded by the supplier.
		secure because of own measures	Expressing the idea that an individual's internet connected devices are secure merely because of the security measures they implemented themselves.
Responsibility of actors	*Responsibilities of user*	user responsible for security (undefined)	Identifying the user or owner of a device as the primary party responsible for ensuring its security.
		user should ensure device configuration	The user or owner of a device having the responsibility to ensure a device is configured in such a way that it is not vulnerable to online threats.
		user should use device safely	The user or owner of a device having to use their device(s) in a responsible manner, for example by not clicking on suspicious links and not installing unknown applications.
	Responsibilities of suppliers	ISP shoud support user on security	The internet provider has a responsibility in helping end-users secure their internet connected devices against online threats.
		devices should be secure out of the box	The belief that internet connected devices should be adequately secure against online threats using the manufacturer's default configuration or security options.
		manufacturers should support user on security	The belief that device manufacturers should support the user in securing their devices by enforcing good security practices or rolling out security updates.
Barriers to centralised security	*Trust*	trust in provider	The degree to which an end-user beliefs that the services provided by their ISP will not harm the end-users.
		trust in other suppliers	The degree to which an end-user beliefs that parties on than the user and provider of a service can be trusted not to harm the interests of the end-user.
	Privacy	implications depend on how it works	The idea that privacy implications may or may not be associated with a service depending on how exactly the service is provided to the end-user.
		lack of transparency about data processing	Insufficient provision of detailed information about the manner in which data is processed; which parties have access to it or what features are used or stored.
		implications are unavoidable or irrelevant	An individual's conviction that potential privacy implications are irrelevant, for example due to them being either unavoidable side-effects of security.
	Costs	financial cost of security	Financial costs or investments expected or known to be associated with the use of security measures, such as purchasing or licensing costs.
		negative effects on performance or productivity	The belief that enabling the malwarefilter might negatively affect productivity or device performance by blocking legitimate internet activity.
	Added value	difficult to assess added value of the malwarefilter	Expressing difficulty in assessing the added value of the malwarefilter, possibly in relation to other services, due to a lack of information or knowledge.
		own measures provide sufficient protection	The belief that the adequacy of an end-users own security measures eliminate the need for additional security measures.
		limited efficacy of the malwarefilter	Limitations to the efficacy of the malwarefilter, for example because of a perception that it can be circumvented by attackers.
Motivating the use of centralised security	*Influence*	malwarefilter was recommended by authority	Influence exerted directly or indirectly by an authoritative party such as the internet provider recommending the use of a specific service.
		malwarefilter provides a feeling of safety	A feeling of enhanced safety obtained by enabling the malwarefilter.
	Ease of use	malwarefilter has management advantages	Advantags associated with the centralisation of related service at a single responsible party, such as internet services and services used to protect device while using the internet.
	Protective ability	malwarefilter provides an extra layer of security	Extra protection afforded by the use of centralised security services in addition to an end-users own security measures or habits.
		malwarefilter prevents malicious activity	The ability of a security measure to prevent malicious activity from occurring.

H

Individual DNS activity



(a) User 1

(b) User 2

(c) User 3

(d) User 4

(e) User 5

(f) User 6

(g) User 7

(h) User 8

(i) User 9

(j) User 10

(k) User 11

(l) User 12

(m) User 13

(n) User 14

(o) User 15

(p) User 16

(q) User 17

(r) User 18

(s) User 19

(t) User 20

(u) User 21

(v) User 22

(w) User 23

Figure H.1: Individual DNS activity patterns of the monitored users. Data could not be retrieved for 2 users.