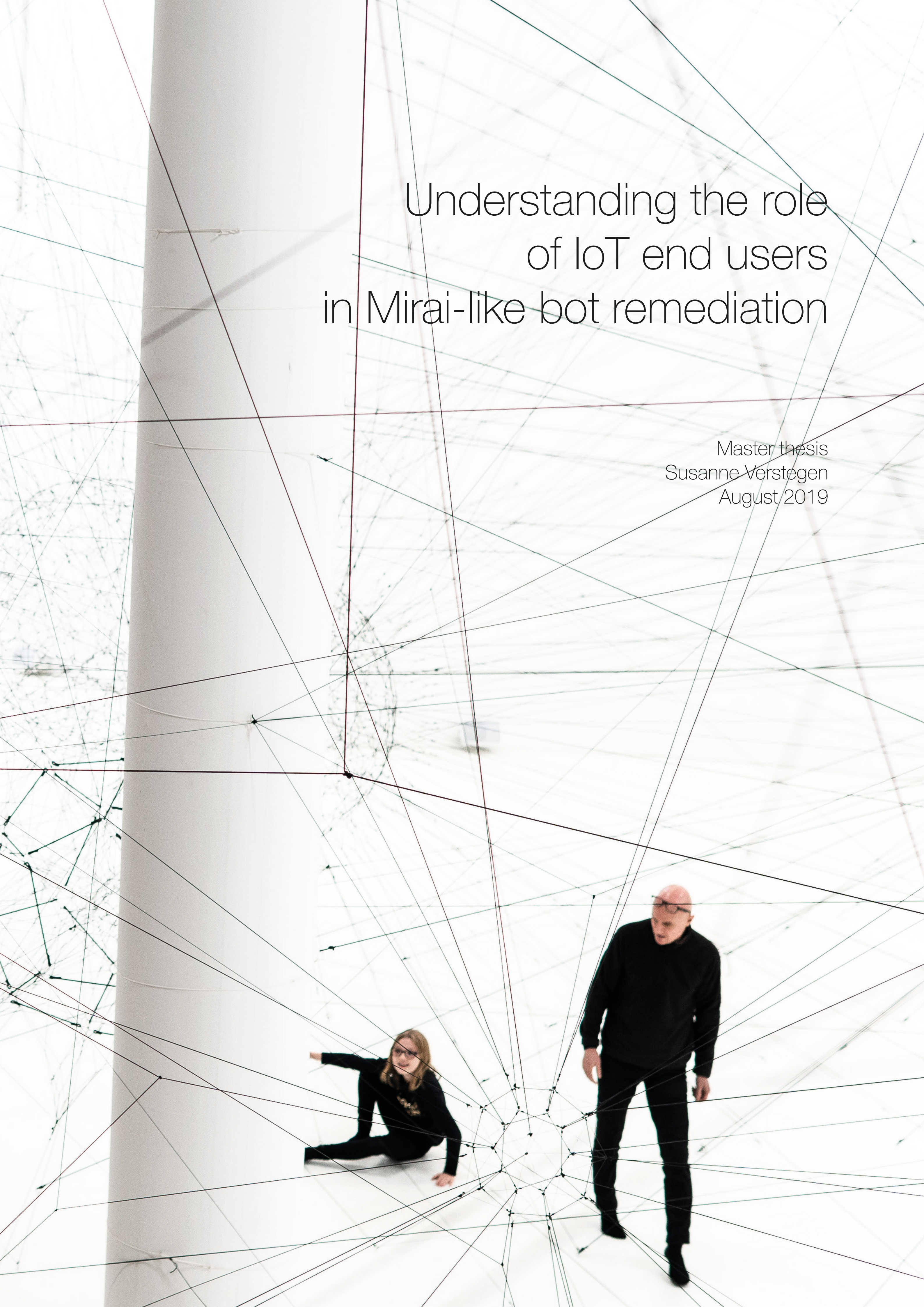
Understanding the role of IoT end users in Mirai-like bot remediation

Master thesis
Susanne Verstegen
August 2019



Understanding the role of IoT end users in Mirai-like bot remediation

Master thesis submitted to Delft University of Technology
in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE
in Engineering and Policy Analysis

Faculty of Technology, Policy and Management
to be defended publicly on August 20, 2019

by

Susanne Verstegen
Student number: 4226453

Graduation committee:

Chairperson	Prof. M.J.G. (Michel) van Eeten, section Organisation & Governance
First supervisor	Dr. C.H. (Carlos) Gañán, section Organisation & Governance
Second supervisor	Dr. S.W. (Scott) Cunningham, section Policy Analysis
External supervisor	D.W.J. (Dennis) van Beusekom, Abuse Desk KPN
Internal supervisor	MSc. E.R. (Elsa) Turcios Rodríguez, section Organisation & Governance
Internal supervisor	MSc. A. (Arman) Noroozian, section Organisation & Governance

This page was intentionally left blank

Acknowledgments

This master thesis marks the end of a personal era. Although graduation is an individual matter, this research would not have been possible with the help and support of several people.

First of all, I'm very grateful to be taken on board in such an interesting and relevant field of research. I want to thank Carlos Gañán as a daily supervisor, who guided me through the world of Mirai and that of academic research in general. Furthermore, I am grateful to be supervised by Michel van Eeten, authority in the field of governance of cyber security, and Scott Cunningham, an expert in policy analysis and data analytics and nowadays professor at the University of Strathclyde. In addition, I want to thank Elsa Turcios Rodríguez and Arman Noroozian for their involvement in the research. Without their critical notes, the research wouldn't have reached the quality level as it does now.

Although I intended to leave my private life out of this preface, I can't escape from mentioning my gratitude for the support I received from my home front. Thank you dad and Jet for your care all these months and thank you mom for your encouraging words during stressful times. Thank you Jesse for lending your great brain whenever I needed a different view on research issues.

I would like to give special thanks to the KPN Abuse Desk. Without them, this experiment wouldn't have been possible in the first place. But more importantly: I am very grateful to them for letting me be part of the team. Graduating can be stressful, but they kept me sane. I want to thank Elmo and Bas for the lunchtime walks, Virgil and Frans for making me laugh, Raymond and Lisette for the always pleasant small talks and Dennis van Drunen for his very motivational words. I want to thank Dennis van Beusekom for guiding me and making sure that I felt home in the SOC. I received indispensable help during the experiment and valuable feedback from everyone in the Abuse Desk. In addition, I want to thank my KPN partner in crime Daniel for sharing so much of his knowledge with me, and being there when I needed help.

Susanne Verstegen

Delft, August 2019

Summary

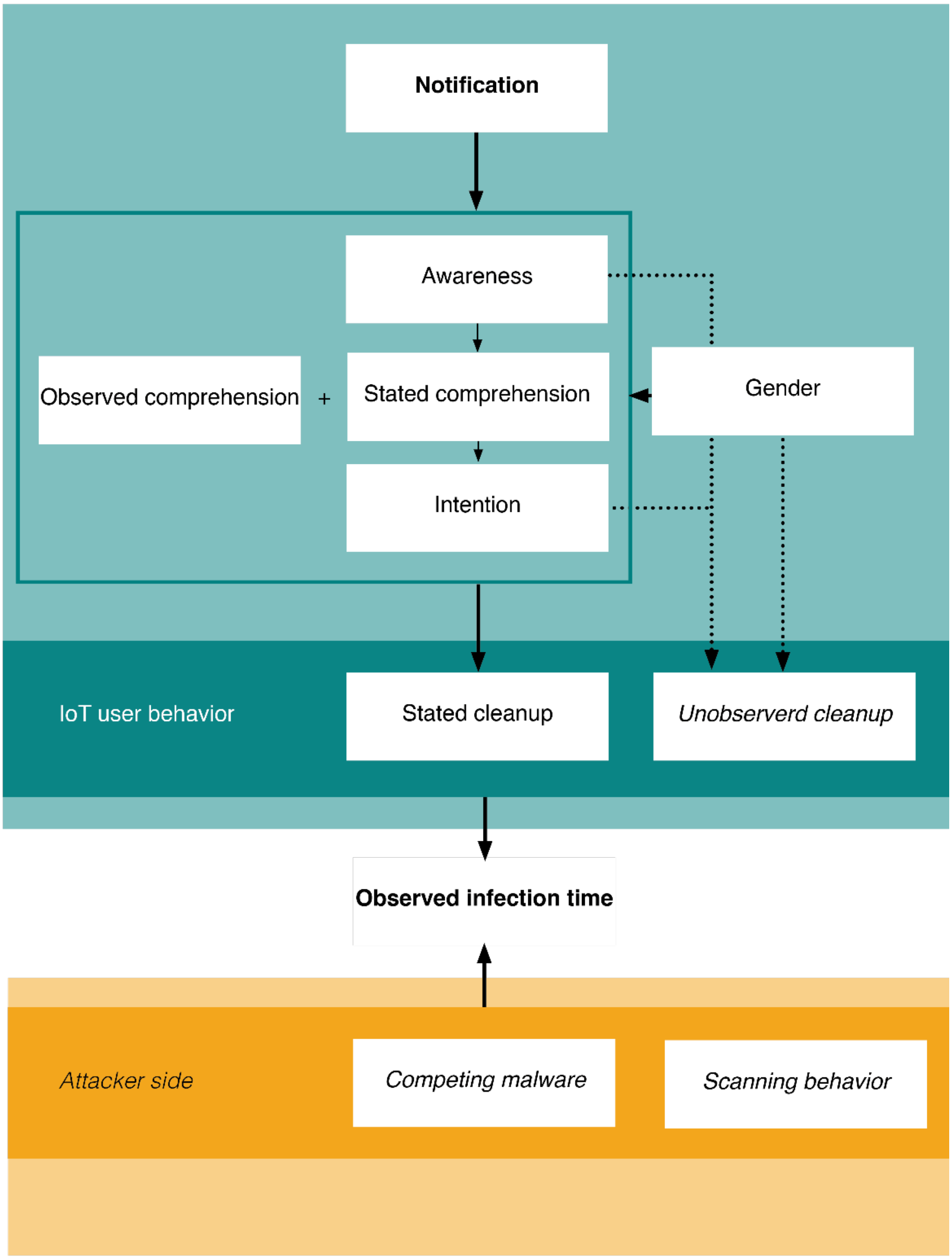
Malicious software such as botnets are a threat to society and increasingly so through Internet of Things (IoT) devices. The large volume, pervasiveness and high vulnerability of IoT devices make them low hanging fruit for malicious actors. Currently, the biggest threat for insecure IoT devices is Mirai, a botnet which is deployed for DDoS attacks. Home users often fail to detect and resolve Mirai on their IoT devices. For this reason, Internet Service Providers (ISP) increasingly take efforts to increase remediation. Sending their infected customers a notifications containing cleanup instructions is currently the most feasible measure on a large scale. However, previous studies point out that it is not clear how people process these notifications, if they comply with it and how this effects the remediation rate and speed.

The central research question of this study is 'What is the role of IoT device end users in Mirai-like bot remediation?'. We have conducted an eight-week experiment at the KPN Abuse Desk that notifies customers about abuse incidents. 177 Mirai-infected consumers have been randomly assigned to a walled garden notification (i.e., a quarantined environment), an e-mail notification, or control group. All subjects within the experiment have been tracked for two weeks to estimate the infection time and are contacted afterward for interview purposes.

Male consumers and consumers younger than 54 years possess relatively more often a Mirai-infected device compared to other consumers. Both e-mail and walled garden notifications are effective in reaching consumers, informing them and encouraging them to take action. The majority of consumers do not follow the recommendations provided by the notification. In contrast, the number of actions that are performed while not mentioned in the notifications is remarkably high. Since many consumers asked for additional help, we conclude that consumers appear don't have a full understanding of how to tackle the problem. In the control group, several consumers remediated Mirai unintentionally. However, these cases do not explain all observed remediation.

Using two survival analysis modeling techniques, we find that consumers placed in a walled garden have a 29% to 85% shorter infection time than other consumers. We conclude that there is a discrepancy between *stated* behavior and the *actual* behavior of consumers. Although we cannot observe all cleanup efforts of consumers, we observed that awareness of the Mirai-infection and the intention to comply with the recommended actions influence that unobserved behavior. Gender also influences the unobserved behavior. Women clean up their device quicker than men while their statements during the interviews contradict this. One explanation is that women may unintentionally clean up their device. We conclude that age, consumer market, device type and customer satisfaction have no significant influence on remediation.

We believe that it is unlikely that all unexplained remediation can be attributed to the unobserved behavior. We thus cannot explain all observed remediation from the user perspective. Therefore, we argue that future work must also focus on the attacker perspective. Since we only observed Mirai-infections, we cannot exclude the possibility that competing malware confiscated infected devices within our experiment. In addition, novel Mirai variants may have evolved scanning behavior which obstructed proper detection of infected bots.


Notification
Awareness
Observed comprehension
+
Stated comprehension
Gender
Intention
IoT user behavior
Stated cleanup
Unobserverd cleanup
Observed infection time
Attacker side
Competing malware
Scanning behavior

Content

List of figures

List of tables

List of abbreviations

AFT	Accelerated Failure Time
AIC	Akaike Information Criterion
CISO	Chief Information Security Office
C&C	Command and Control Center
C-HIP	Communication-Human Information Processing
CI	Confidence Interval
DMZ	Demilitarized Zone
DNS	Domain Name System
DDoS	Distributed Denial of Service
ENISA	European Union Agency for Network and Information Security
GCA	General Cyber Alliance
HR	Hazard Rate
IoT	Internet of Things
IP	Internet Protocol
IPR	Interview Protocol Refinement Framework
IQR	Interquartile Range
ISP	Internet Service Provider
LL	Log-Likelihood
LRS	Likelihood-Ratio Statistic
PMT	Protection Motivation Theory
RCT	Rational Choice Theory
SOC	Security Operation Center
TPB	Theory of Planned Behaviour
UPnP	Universal Plug and Play

1 Introduction

1.1 Background

1.1.1 Internet connects

The Internet connects beyond people: it increasingly connects ‘things’. We find ourselves at the start of this new paradigm called the Internet of Things (IoT). The term refers to the concept of interconnected objects which can send data to other objects, systems, and people. There is no definition of IoT that is widely accepted: some definitions focus on the architectural requirements of an IoT environment, whereas others emphasize the ubiquity and autonomy of IoT networks. A unique characteristic of IoT is so-called ‘smartness’ of its networks: each object is connected to a network (to gain access to the Internet or share data with other devices), is context-aware (the device perceives information from its environment) and is autonomous (can perform tasks without the user’s command) (Silverio-Fernández, Renukappa, & Suresh, 2018). Collectively, these smart things have the capability to ‘collect, process and exchange data [in a network] in order to adapt dynamically to a context’ (ENISA, 2017).

IoT applications are in numerous places. One example is smart homes. A smart thermostat can be switched on or off from outside the house via an app. A smart smoke detector can check the functionality of its sensor or can send a warning to your mobile phone when it detects smoke. These and other IoT devices such as smart lights, fridges, camera’s and faucets can make daily life activities easier, safer or greener (Essent, 2017). Other IoT environments can be found in transport (smart public transport, smart airports, smart cars), health (eHealth, smart hospitals) and overarching infrastructures (smart grid, smart cities).

1.1.2 Botnet of things

While the number of Internet-connected devices grows, so does the concern regarding their security. The European Union Agency for Network and Information Security (ENISA) identifies twelve generic issues that impede the secure use of IoT (ENISA, 2017). We provide three examples to give an impression of the obstacles:

- Limited device resources: most conventional security controls cannot be adopted by IoT devices due to technical constraints such as low computing power;
- Insecure programming: due to a short ‘time to market’ and slow adaption of guidelines and regulations, vendors give low priority to security and privacy of their devices;
- Absence of user interface: most IoT devices do not have a user interface, which makes it more complicated for users to perform security measures such as changing the default password or performing updates.

In short: many IoT devices are not secure, or they are used in an insecure manner and thereby form the risk of being abused by malicious actors. IoT devices can be abused in several ways, but the most urgent threat is that of weaponization through so-called botnets (derived from ‘robot networks’). A botnet is a coordinated network of compromised hosts[1] infected with malware which is controlled by a malicious actor without the owner’s knowledge (International Telecommunications Union, 2008; Livingood, Mody, & O’Reirdan, 2012). These networks collectively increase the computing power and bandwidth of a criminal which can be used to serve different criminal activities such as generating spam e-mails, launching Distributed Denial of Service (DDoS) attacks, destruction of data, identity theft and click fraud (International Telecommunications Union, 2008; Livingood et al., 2012). A DDoS attack is an attempt to flood a target with Internet traffic by means of a magnitude of compromised systems. The target, often a server or network, can get overwhelmed which results in a disruption of its services.

In 2016, one Mirai botnet compromised more than 600K IoT devices and overwhelmed the world by DDoS-attacks of high profile targets such as a Domain Name System (DNS) infrastructure (Antonakakis et al., 2017; Groenewegen, 2016). The Mirai botnet shows the destabilizing potential of botnets and the danger of poor security of low-end devices. ENISA (2019) points out that malware authors increasingly target IoT devices and that the trend of botnet attacks is increasing. This trend goes hand in hand with the growing number of IoT devices and the increasing range of their application. Predictions about the number of installed IoT devices in 2020 vary between 20 and 50 billion (ENISA, 2017; Statista, 2019).

1.1.3 Voluntary compliance

Internet end users appear to struggle to detect and clean up Mirai on the IoT devices they use (Orçun Çetin, Altena, Gañán, & Eeten, 2018). Internet Service Providers (ISP) are in the unique position to stimulate malware remediation because they have the capabilities to detect malicious activities in their network and are able to identify and thus notify the infected end user (Livingood et al., 2012). For this reason, ISPs are often a designated actor to make botnet mitigation efforts (Orçun Çetin et al., 2018; Livingood et al., 2012). Besides their natural control position, ISPs also have an incentive to mitigate botnets due to the increasing costs and reputational damage they suffer from their polluted network. In the last decennium, industrial collaborations, governments and academics have published best practices, recommendations and studies about botnet mitigation in ISP networks.

ISPs can thus deploy mitigation measures against Mirai by notifying infected consumers. There are two typical notification mechanisms. The first is placing an infected customer into a quarantined environment, a so-called ‘walled garden’, and instructing them what to do to clean up the infection. This measure prevents the Mirai botnet of extending and draws a customer’s attention to the recommended actions. The second option is only to warn the customer without further consequences and provide him or her with cleanup instructions. Both measures appeal to an end user’s willingness to comply with the instructed remediation advice although the first (walled garden) is more intrusive due to the disconnection from the Internet. Also, walled gardens raise major objections from customers and are time- and cost consuming for an ISP (Orçun Çetin et al., 2019).

1.1.4 Prior research

Although the effectiveness of abuse and vulnerability notifications is broadly studied, the amount of work focusing on IoT abuse is small. The master thesis research by Lisette Altena (2018) and the subsequent articles by Çetin et al. (2018, 2019) were ground-breaking and hitherto the only empirical studies of IoT malware cleanup in the wild.

[1] A ‘host’ refers to a computing device that is connected to the Internet (Livingood et al., 2012)

These studies conclude that walled gardens are effective in terms of clean up rate and speed and that e-mail only notifications do not have more impact on remediation compared to the control group that did not receive a notification. Striking is the high natural remediation rate of the control groups, which was 77% after 14 days (Altena, 2018, p.53). The high natural remediation is partly attributed to the non-persistent character of Mirai, which means the malware is remediated after a device is switched off (Orçun Çetin et al., 2019). These results raise questions about what underlying behavior of end users cause these findings. It is unclear if and why notified users do not comply or whether they fail to perform the recommended cleanup actions.

1.1.5 A socio-technical domain

The system under study is sociotechnical in nature. The IoT paradigm and its negative side effects such as Mirai emerge from an interaction between the social and technical domain. The technical aspects in the system lie in the increasing capabilities and applications of IoT devices. At the same time, the sophistication of the abuse of these devices is also a technical component of the problem. This sophistication will be concretized in the context of Mirai in section 2.2.

The social aspect of the system interacts with this technical domain: due to the increasing functionalities and better access to IoT devices, more people buy and use IoT devices. In addition, the security of an IoT device is not solely determined by its design; how a person configures and uses a device is of great influence for its exposure to potential abuse. The development within the technical domain thus stimulates the presence of IoT in the social domain and vice verse. The size of the IoT paradigm also increases the scale of the negative consequences of insecurity. The problems of insecure IoT devices are tangible on society-level: for example companies, Internet service providers and governments suffer from the consequences of insecure IoT devices through DDoS attacks.

This interaction between the technical and social domain of IoT devices and their (in)security create a complex system in which both aspects cannot be considered in isolation. One must understand both domains and interaction between them to understand the problems that emerge from this system and how to mitigate them.

1.2 Research objective

1.2.1 Problem statements

Notifications are currently one of the most feasible mitigation measures to fight IoT abuse on a large scale. Three problems can be identified that hold back good functioning remediation efforts:

A) Altena’s (2018) research provides the first empirical findings on remediation rates and speed of different notification mechanisms. There is no empirical explanation yet for the high remediation rate in the control group. First, we want to find out to what extent remediation can be explained as a result of notifications. Also, we want to explore what factors may explain remediation among consumers that are not notified. Then, to improve the effectiveness of notifications, we must understand why customers (do not) comply and why they (do not) perform the recommended actions.

B) The current bottleneck for an ISP is workload capacity. The more infected customers are sent a notification or are placed in a walled garden, the higher the workload for an ISP since customers regularly e-mail the Abuse Desk with questions (Altena, 2018). Responding and providing assistance to customers is a process that requires personal dedication and is difficult to automate. To decrease questions, we need to understand how customers perceive the content of a notification, if this aligns with the intended message, and whether customers understand the intended message.

C) An ISP's key service is Internet access. Walled gardens are not a preferable solution on a large scale since it disturbs this key service. Also, putting customers in a quarantined environment requires computing capacity. However, the other alternative – e-mail only notifications - have a lower remediation rate and are thus less effective. This creates a trade-off between inconvenience and effectiveness of a notification. To achieve more customer-friendly notifications, we must understand when customers are dissatisfied and how they wish to be approached.

1.2.2 Research questions

The problems discussed in the previous section suggest a need for a better understanding of what drives remediation. Due to time and technical constraints, this research will particularly focus on remediation of the *Mirai malware* by *home users* (consumer market). As will be discussed in chapter 2, Mirai is currently the most serious and predominant form of IoT abuse. Home customers – in contrast to business customers – are easier to reach and notify since the contact details are that of the person of interest. To achieve the defined objectives, the research question is as follows:

(RQ) What is the role of IoT device end users in Mirai-like bot remediation?

The following five sub-questions (SQ) break down the main research question formulated above into smaller questions which need to be answered:

(SQ 1) What are the characteristics of IoT device end users who get Mirai-infected?

The first step in exploring the role of end users that are Mirai-infected is examining who these consumers are. We will explore the age and gender of infected IoT users in a real-life setting ('in the wild'). This exploration is executed for two separate populations: consumers of the ISPs KPN and Telfort. Telfort is a budget subsidiary of KPN.

(SQ 2) What actions do Mirai-infected consumers perform?

Notified consumers are asked to perform a set of recommended actions to clean up their infected device. It is yet unclear how many consumers that have the intention to comply with the notification, succeed in doing so correctly. We want to know which actions consumers perform and to what extent this influences remediation. It is also not empirically explored yet what actions non-notified customers perform (intentionally or unintentionally) that cause remediation (Altena, 2018).

(SQ 3) What are the reasons for non-compliance with Mirai notifications?

One can think of different reasons for non-compliance with the recommended cleanup actions. Notification delivery failure, misunderstanding and a lack of motivation are three examples. To be able to encourage consumers to take voluntary action, one must know the common obstacles that stand in the way of compliance. A large body of literature provides models and empirical findings of these reasons in the context of security and vulnerability notifications

(Orçun Çetin et al., 2019). There is no study yet that explores this in the case of IoT abuse notifications.

(SQ 4) How do consumers experience Mirai notifications?

To improve future notification effectiveness, we must also consider the notification experience of notified customers. To this end, we want to explore their opinions and suggestions.

(SQ 5) How can remediation of Mirai-like bots be explained?

The knowledge gaps that come forward in Altena's (2018) research are a direct motivation for the existence of this study. This research will further explore the effect of notifications on remediation and will do so by replicating the experimental setup of previous research. In addition, this study will make use of statistical data modeling techniques to explore the effect of notifications and other factors on remediation.

1.3 Research approach

To answer the research questions, we use a mixed-methods approach. This kind of design involves 'collecting, analyzing and interpreting both quantitative and qualitative data'. The core assumption of this approach is that the combination of both forms of data will provide a 'more complete understanding of a research problem than either approach alone' (Creswell, 2014, p.4). There are different typologies of mixed methods; this design is *convergent* and *parallel*. *Convergent* mixed methods are a form in which qualitative and quantitative data is converged or merged to create a comprehensive view of the research problem (Creswell, 2014). *Parallel* refers to the data collection sequence: both forms of data will be collected in parallel and the results will be integrated during their interpretation (Creswell, 2014).

The study exists of four phases, which will be discussed in the next sections. The research will be executed in cooperation with the Abuse Desk of KPN, a Dutch ISP. The Abuse Desk mitigates abuse incidents among KPN and Telfort customers (KPN, n.d.). Telfort is a budget subsidiary of KPN. Chapter 3 provides more information about KPN and its network abuse mitigation practices.

1.3.1 Research context

The first phase provides the context in which the research will be executed. A literature review will explore studies into cybersecurity behavior, notification effectiveness and IoT security challenges. This chapter provides us with a solid background to base the study on.

To set up an experiment and to understand the context in which a customer deals with a notification, we must understand how KPN detects and notifies infected customers. These practices by the Abuse Desk are studied by observing the variety of activities and by engaging with the employees of the Abuse Desk. The reporting of the practices will be validated through reviews by two Abuse Desk employees.

1.3.2 Experiment

Over eight weeks, a randomized controlled experiment is performed to explore how customers deal with different notification mechanisms. All IoT-infected customers that appear on the Abuse Desk's radar within this time frame are included in the experiment. After random assignment to a group, each customer is tracked for two weeks to measure the infection time of the bot. After tracking, the customer will be contacted to perform an interview in a semi-structured manner. More details about the experiment set up and limitations are presented in chapter 4.

1.3.3 Data analysis

Quantitive data concerning remediation speed and rate and qualitative data concerning consumers' characteristics, behavior and reaction have been collected during the experiment. This data will be collectively analyzed to answer the six research questions using exploratory modeling. The methods used for this are supported and described in chapter 4.

1.3.4 Research evaluation

The last phase entails the evaluation of the research. Using the results and conclusions of the five sub-research questions, the central question will be answered. The research quality will further be assessed by discussing the limitations of the study and the validity of the results. Lastly, we will provide several suggestions for future research.

1.4 Academic & societal relevance

1.4.1 Societal relevance

The combat against IoT abuse is not only in the hands of ISPs. Therefore, insights of this study regarding Mirai and the role of an infected device owner may help to provide a better understanding for a greater range of stakeholders. Policy decisions regarding botnet mitigation and IoT security in general, may benefit from such insights as well other Mirai botnet mitigation actors such as other ISPs.

The Dutch government is increasingly more aware of the damaging consequences of poor IoT device security which is reflected in increased budget and measures (Ministerie van Economische Zaken en Klimaat, 2019; Raad Cyber Security, 2017). One of these measures is an awareness campaign that will start in October 2019 'aimed at changing behavior [of citizens and enterprises]' (Ministerie van Economische Zaken en Klimaat, 2019). Findings of this study will be of added value for such purposes because they A) help to understand the characteristics of the target audience, and B) provide insight in the troubles that IoT end users perceive when cleaning up an infected IoT device. Both findings will increase the effectiveness of an awareness campaign as proposed.

In addition, there is a need for more certainty regarding remediation. Although the study of Altena (2018) shows positive results regarding the use of walled gardens, remediation among unnotified consumers cannot be explained. This is problematic in the context of policymaking since either we don't understand the cause of remediation, or our monitoring instrument is not reliable and give us a distorted picture of Mirai remediation.

The focus of the experiment is on Mirai remediation in the Netherlands since the study population exists of KPN and Telfort consumers. However, the findings of this study are of added value across borders. For example at the EU level, the ENISA is increasing its efforts to address IoT safety and security challenges (ENISA, 2017). Raising awareness is one of the baseline security recommendations which may also benefit from best practices in the Netherlands, including the findings of this study.

1.4.2 Academic relevance

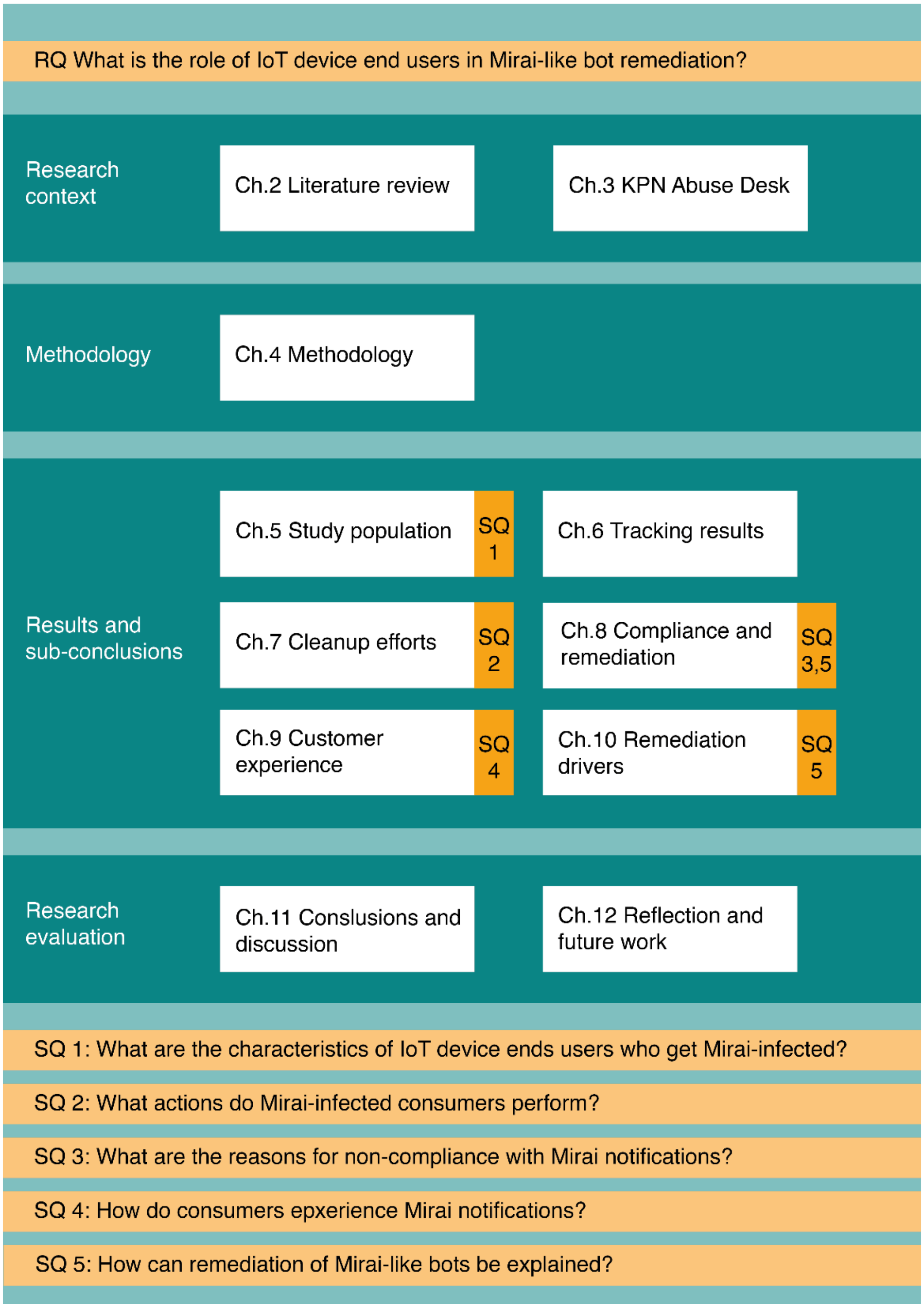
As addressed before, notification effectiveness of IoT abuse notification is a recent terrain in academia. Altena's (2018) research provided novel insights in this area and thereby simultaneously created new knowledge gaps in how to explain the observed 'natural' remediation of Mirai. This research attempts to fill these gaps by looking into the role of IoT users.

1.4.3 Added value KPN / other ISPs

Since the treatments in the experiment are equal to KPN's common practice, we can obtain reliable results of how KPN and Telfort customers deal with and perceive Mirai notifications. This understanding may not only help to improve notification effectiveness, but also customer satisfaction. Improved notification effectiveness can reflect in time-saving among Abuse Desk employees in helping customers and overall improved costs efficiency. In addition, other ISPs may benefit from the best practices of KPN's Mirai remediation efforts and the insights provided by this study.

1.5 Thesis organization

The organization of this thesis report is schematically illustrated in figure 1. The research context will be described in chapter 2 (literature review) and chapter 3 (KPN Abuse Desk). The experimental setup and statistical tests that will be used, are presented in chapter 4 which covers the research methodology. Chapter 5 explores the study population and thereby provides answers to research question 1. Chapter 6 presents the tracking results. This chapter does not answer a research question but rather provides a general view of the data which serves as a base for the following chapters. Chapters 7 to 9 each cover results and sub-conclusions of the research questions 2 to 4. The analysis in these chapters is mainly of qualitative of nature. Chapter 10 combines all data by modeling and thereby answers research question 5 (which is already partially answered in chapter 8). In chapter 11, the overall key findings are recapped and the main conclusions are drawn, which provide an answer to the main overarching research question. Reflection upon the research design, results and conclusions are presented in chapter 12.



RQ What is the role of IoT device end users in Mirai-like bot remediation?

SQ 1: What are the characteristics of IoT device ends users who get Mirai-infected?

SQ 2: What actions do Mirai-infected consumers perform?

SQ 3: What are the reasons for non-compliance with Mirai notifications?

SQ 4: How do consumers epxerience Mirai notifications?

SQ 5: How can remediation of Mirai-like bots be explained?

Figure 1 Thesis organization

2 Literature review

2.1 Introduction

This chapter provides a summary and evaluation of works that are related to the research problems in question. We can distinguish three conceptual categories which are of interest to understand customer behavior after abuse notifications concerning Mirai:

IoT and the emergence of Mirai

The emergent IoT paradigm forces a change of security thinking and practices. IoT abuse practices such as the Mirai malware bring up issues on how to overcome the poor security of devices and insecure consumer behavior. But how is this different to abuse of conventional devices? Why is Mirai such a threat? And how does Mirai operate?

Notification effectiveness

Within the field of cybersecurity, warning and vulnerability notifications are a broadly studied topic. These studies often focus on the effectiveness of a notification - to what extent a notification leads to the desired outcome. How can we define notification effectiveness? And what are the best practices to achieve effective notifications?

User cybersecurity behavior

The vulnerability of a device – and with that, the risk of abuse - is partly determined by the behavior of its user. Think of setting safe passwords, regular updates, non-clicking on suspicious links, etc. Theories and models from a variety of disciplines provide different explanations of cyber (in)secure behavior of users. Why do users comply to abuse-notifications from the perspective of these works? What may explain non-compliance?

The literature search is conducted through the framework as proposed by Webster and Watson (2002) and Levy and Ellis (2006). They propose the following three steps in identifying relevant literature:

- Keyword search: the initial step using keywords in scholarly databases and leading journals;
- Backward search: reviewing citations of (relevant) articles;
- Forward search: use academic search engines to find articles that have cited the (relevant) articles.

This initial literature search resulted in 124 articles and books that are structured by concepts and relevance (three-point scale). Appendix A provides the details on the literature search, such as used keywords and search engines. Appendix B provides a table with takeaways from the relevant studies from the literature search. The findings of the literature review are structured following the three conceptual categories in sections 2.2 to 2.4.

2.2 IoT and the emergence of Mirai

2.2.1 Low hanging fruit

The Internet of Things is a new kid on the block. A 2011 whitepaper of Cisco argues that while the World Wide Web knows several evolutionary leaps, IoT is the first evolution of the Internet itself (Evans, 2011). The concept of IoT was first used in 1999 in a networked radio frequency identification (RFID) group in the Massachusetts Institute of Technology (Evans, 2011; Heer, René, Loong, Sandeep, & Klaus, 2011). Since then, IoT technology and applications have developed and finds itself at the foundation of the new paradigm of interconnectivity.

In 2008, the number of Internet-connected devices exceeded the total world population (Evans, 2011). The numbers are growing, and prognoses estimate further increase to 20 to 50 billion devices in 2020 (ENISA, 2017; Statista, 2019). This magnitude is an important cause of the threat that stems from these devices: 'What they lack in computational capabilities, they make up in numbers' write Vlajic and Zhou (2018). This trend, in combination with poor security and often unbroken connection to the Internet, makes IoT devices 'low hanging fruit for hackers' (Kolias, Kambourakis, Stavrou, & Voas, 2017).

Whereas security practices have become common practice in traditional devices (laptops, smartphones, etc.), security of IoT devices is a complex matter. This makes IoT currently the 'weakest link in the security chain of computer networks'. The vulnerable character of IoT devices can be explained by an accumulation of reasons:

- In a rush to market, vendors minimize or neglect security to keep costs low, time-to-market short and their devices user-friendly (Kolias et al., 2017; Raad Cyber Security, 2017). On top of that, the security of IoT devices is a difficult task since many of them use lightweight operation systems on which traditional computer security solutions cannot be run (Batalla, Mastorakis, Mavromoustakis, & Pallis, 2017).
- Customers often think of IoT as plug and play devices and want to make sure their device works quickly rather than investigate and taking basic security measures such as setting a new password (Vlajic & Zhou, 2018). This behavior is often strengthened by the (lack of) interfaces of IoT devices, which are non- or minimal interactive (Kolias et al., 2017).
- If security measures are in place, two reasons above contribute to poor maintenance of security: vendors may not develop security patches and users may not think about/forget about updating their device regularly (Bertino & Islam, 2017; Kolias et al., 2017; Vlajic & Zhou, 2018).
- Deployment of global security mechanisms or policy is not possible due to the distributed control of the Internet and complex governance structure. Each network and country follow their local rules and many actors are involved (Donno, Dragoni, Giaretta, & Spognardi, 2018; Raad Cyber Security, 2017). ENISA (2018) concludes that all security requirements for IoT can be met with existing standards, but that a new flexible and holistic approach is needed to actually achieve effective IoT security in a dynamic ecosystem.

2.2.2 Mirai: 'The Future'

Currently, the biggest threat for insecure IoT devices is Mirai, a botnet which is deployed for DDoS attacks. DDoS attacks can have severe destabilizing consequences for the direct victims of an attack as well as for the systems and users that depend on that service. DDoS attacks illustrate the interdependent nature of Internet security: the vulnerability of a DDoS victim is not determined by the security his/her own system, but instead by the security of the entire Internet (Donno et al., 2018).

Although there are more DDoS-capable IoT malware (Donno et al. identify twelve others), Mirai stands out because of the damage it has caused and due to its growing technical sophistication (Donno et al., 2018). The Mirai malware (Japanese for 'the future') was first identified by a whitehat security research group in 2016 (Kolias et al., 2017). In that same year, the Mirai source code was published open source on the online software development platform GitHub, which gave birth to a number variety of variants and imitators, often more sophisticated and with new capabilities (Antonakakis et al., 2017; Donno et al., 2018; Kolias et al., 2017). Collectively these variants are referred to as 'Mirai-like'.

A Mirai botnet has the following four components (Donno et al., 2018; Kolias et al., 2017):

- A 'bot' can be considered as an infected device. Strictly speaking, the term refers to the malware that runs on the device;
- A Command and Control (C&C) center is a server that provides the botherder (the malicious actor that runs a botnet) with an interface to control the botnet;
- The report server receives information about newly infected bots and forwards this to the loader server;
- The loader server uploads the Mirai malware code to the newly infected devices.

Due to the public availability of the source code, the operations of Mirai have been largely studied. The following description of Mirai is based on the articles of Antonakakis et al. (2017), Donno et al. (2018) and Kolias et al. (2017). A Mirai bot functions as follows:

Step 1 The first phase exist of scanning randomly public IPv4 addresses through TCP port 23 and 2323 (Telnet protocol). IP addresses on a hard-coded blacklist that include ones of the U.S. public services and the Internet Assigned Numbers Authority are excluded from this scanning.

Step 2 When a potential victim is identified, the bot will execute a brute-force[2] attack the victim device with ten random username-password combinations from a hard-coded list of 62 credentials.

Step 3 If the brute-force login has succeeded and a Telnet connection is established, Mirai sends the IP address of the victim and the correct credentials to a report server.

Step 4 The report server forwards the information to a loader server, which logs in on the victim device, determines the hardware architecture and downloads and executed the Mirai malware that fits the system.

Step 5 After the successful download and execution of the binary code, the binary code is deleted. The malware is now active and has four tasks:

- Scanning: searching for new victims, see step 1
- Killing: kills other processes bound to TCP/23 and TCP/2323 and prevents breaking in of others through other common methods to protect itself from competing malware (and thereby maximize availability)
- Waiting commands: once in a while, the bot checks-in with the C&C server and waits for further commands.

[2] A brute-force attack is an automated trial-and-error method (i.e., automated 'guessing') used by hackers to obtain encrypted data, often login credentials.

- (DDoS) attacking: when the C&C server gives an attack command, the bot will attack the target server with one of the ten available attack variations.

While TCP/23 and TCP/2323 were initially used to lay a connection, new strains also target other ports. Cetin et al. (2019) observe fourteen target ports, distributed over six protocols. Devices providing HTTP-related services are most frequently compromised by Mirai (Antonakakis et al., 2017; Orçun Çetin et al., 2019). By looking into the credentials that are hard-coded in Mirai, studies by Antinajajus et al. (2017) and Cetin et al. (2019) both find that IP cameras, DVRs, and consumer routers are the most targeted types of devices.

2.2.3 IoT governance

Traditionally, the responsibility of, for example, a DDoS attacks lies with the users of a host: they have a duty of care with regard to maintaining secure devices (Kolias et al., 2017). However, due to the different nature of IoT as covered in section 2.2.1, insecure IoT devices cannot purely be attributed to its end users. In an advisory report to several ministries, the Dutch Cyber Security Council raises concerns regarding the liability and duty of care of IoT devices (2017). 'The IoT playing field is big, borderless and knows a complex international composition. [..] Due to the great number of primary international players on the IT-market, there is a lack of overview' (Raad Cyber Security, 2017). This chaotic situation complicates the question of who is responsible for IoT security.

In addition to the immaturity of the IoT governance ecosystem, the governance of botnet mitigation in general is known for its complex character. In the last decade, many initiatives are taken in so-called stakeholder communities: groups of actors that are related for geographical reasons (EU, Netherlands, etc.) and/or functional reasons (law enforcement agencies, ISPs, etc.). Due to the overlap of these initiatives, the landscape of botnet mitigation is diverse and dispersed (International Telecommunications Union, 2008). Stakeholder communities tend to operate in their silo while meanwhile, overarching coordination is missing (International Telecommunications Union, 2008).

Despite the urgency of the IoT security and IoT botnet mitigation, there is hitherto no governance structure in place that 'glues' all stakeholders and their interest (Almeida & Goh, 2017). Rules, norms and regulation concerning IoT security (both cross-bordering and cross-sectoral) will thus not be implemented in the foreseeable future (Orçun Çetin et al., 2019). Meanwhile, best efforts are made by different stakeholders to mitigate IoT abuse. Notification efforts by the Abuse Desk of KPN are such an example and are said to play a 'critical role' (Orçun Çetin et al., 2019). More information about KPN's notification practices is covered in chapter 3.

2.3 Notification effectiveness

Abuse and vulnerability notifications call for end users' willingness and capability to execute the recommended action voluntary. The success of these notifications is mainly measured through infection tracking. Studies into this focus on different potential predictors of remediation such as user traits, notification content and notification channels. These predictors are discussed in the following sections. Since most studies are performed under usual circumstances (real-life), we speak here of 'effectiveness' rather than 'efficacy' (ideal or selected circumstances). Note: the following findings come from studies into *abuse* notifications as well as studies into *vulnerability* notifications.

2.3.1 Notifying pays off

Li et al. (2016) and Çetin et al. (2017) both analyze which aspects of vulnerability notifications lead to higher remediation rates. Both studies show a higher remediation rate when notifications are sent but lack clear insights into the incentives that have led to this remediation. Vasek & Moore (2012) find that sending more than one notification does not create a higher remediation rate. The most recent studies by Altena (2018) and Çetin et al. (2018, 2019) show a high natural remediation rate of Mirai and a very low reinfection rate, as discussed in section 1.1.4. The low reinfection rate creates a discrepancy with earlier lab results, which is not well understood.

2.3.2 Notification content

Krol et al. (2012) explore whether computer users heed warnings and conclude that the majority of people ignores security warnings and that the content of the warning does not matter. In contrast, Vasek & Moore (2012) conclude that for abuse reporting, detailed descriptions of a compromise lead to a higher remediation rate. This is also supported by an empirical study of Çetin et al. (2016) into the role of sender reputation. Whereas information on the compromise must be detailed, Forget et al. (2016) argue that security instructions, on the other hand, must be very simple. Stock et al. (2018) find a discrepancy between problem awareness and actual patching efforts and therefore claim that content of notifications is key to convince the receiver.

2.3.3 Notification channel

The commonly used notification channel to reach end users is e-mail. Çetin et al. (2016) conclude that sender reputation does not matter. Stock et al. (2018) argue that e-mail as a communication medium suffers from several shortcomings but that other channels do not justify their significant financial costs and time overheads. This is contradicted by the studies by Çetin et al. (2018, 2019) who observed that among consumers placed in a walled garden, 92% of the Mirai infections is remediated after two weeks. Although this measure is highly effective, fifteen percent of the customers expressed dissatisfaction with this intrusive measure and the solution is not cost-effective on a large scale. The two articles also conclude that e-mail notifications did not have an impact on remediation compared with the control group (respectively 77% and 74% infections were remediated after two weeks).

2.3.4 User traits

Interestingly, Krol et al. (2012) conclude that people with a lack of computer knowledge revealed saver computer behavior and that participants rely on their own judgment, rather than a security warning. Forget et al. (2016) compares user engagement (‘desire to control and manage their computer’s functionality and security’) with the actual security state of their computer and concludes that user engagement alone is not a good predictor for computer security. This implies that even with the right motivation, people’s behavior may not result in the desired outcome.

In conclusion, although notifications in most studies lead to a higher remediation rate, the results are quite modest. Several articles argue that content is a driving factor, but there is not yet a universal understanding of the criteria of a successful notification. Besides an incomplete understanding of the influence of content, differences among end users may also affect the success rate of notifications. A one-size-fits-all notification is, therefore, an illusion.

2.4 Cybersecurity behavior

To encourage end users to take voluntary action, one must understand their motives. Studies on this topic often use a theory or model to illustrate the antecedents or drivers that explain (in)action. Within the consulted literature, fifteen of these models and theories are identified. Appendix C presents these behavioral models, their explanatory value and whether or not they are relevant in light of IoT abuse remediation. Thirteen of those theories have psychological fundamentals; one stems from the warning science and one from economics. This section discusses four theories that proved to be valid in understanding why and when people do not comply with security measures recommended by notifications.

2.4.1 The Theory of Planned Behaviour

The Theory of Planned Behaviour (TPB) is a psychological model that is often used in explaining how individual security behavior is influenced. Ajzen (1991) proposes the TPB to predict actions based on an individual's *intention* to perform that behavior. Intentions are 'indicators of how hard people are willing to try' to perform a certain behavior and is influenced by three 'motivational factors' (Ajzen, 1991, p.181):

Attitude toward the behavior: the users' positive or negative feeling towards engaging in a particular behavior (Ifinedo, 2012a; Safa et al., 2015).

Subjective norm: 'the perceived social pressure to perform or not to perform the behavior.' (Ajzen, 1991, p.188)

Perceived behavioral control: the 'perceived ease or difficulty of performing the behavior and it is assumed to reflect past experience as well as anticipated impediments and obstacles.' (Ajzen, 1991, p.188). Perceived behavioral control not only predicts intention, but also influences actual behavior directly (Ajzen, 1991; Howe, Ray, Roberts, Urbanska, & Byrne, 2012). This variable is recognized by Ajzen (1991) to be congruent with the notion of 'self-efficacy' as introduced by Bandura in 1977.

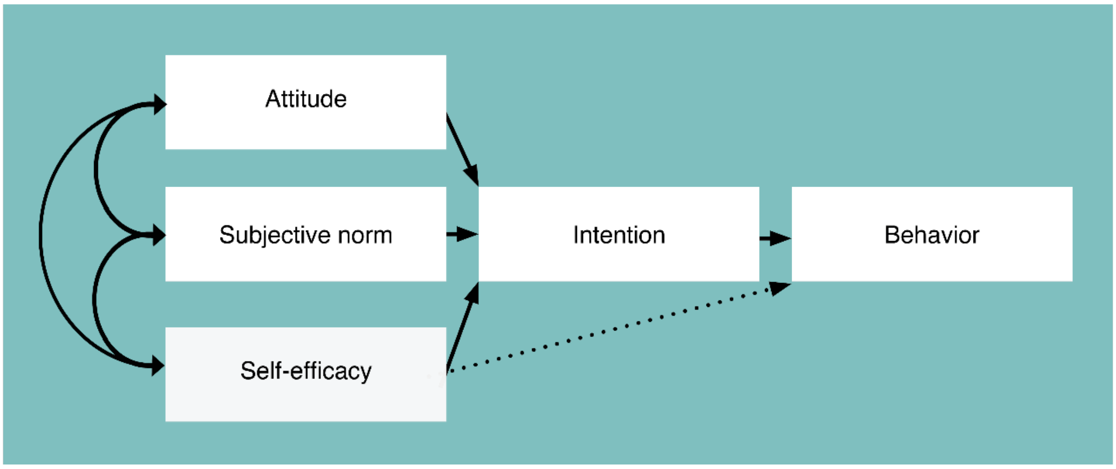


Figure 2 Theory of Planned Behaviour. From Ajzen (1991)

The relations between the motivational factors are illustrated in the model in figure 2. Although the model is quite general, it provides several insights that are useful in the context of IoT abuse remediation. Firstly, *intention* is the most important predictor for behavior, which means that intention does not always lead to the desired behavior. This intention-behavior discrepancy also comes forward in other psychological models such as TPB's predecessor Theory of Reasoned Action (TRA) and the protection motivation theory (PMT, see section 2.4.3). Sheeran

(2002) quantified this gap through a meta-analysis. He found that intention explains 28 percent of the variance in future behavior. A second insight is that self-efficacy is assumed to be of direct influence on behavior. This has two reasons: someone with more self-confidence in achieving something is more inclined to make more effort, and perceived control is a good measure for actual control (Ajzen, 1991). If people thus believe they could perform a certain action, the probability that they succeed is higher. A last valuable insight is that *attitude toward the behavior* appeared in some studies to be influenced by knowledge since knowledge creates more awareness (Dinev & Hu, 2007; Safa et al., 2015). One can thus assume that better information provision can lead to more desired behavior.

2.4.2 Rational Choice Theory

The Rational Choice Theory (RCT) is a neo-classical approach to understand behavior and crossed the boundaries of the economics domain due to its explanatory power. Like the theory of planned behavior, the RCT focuses on the determents of behavior. The theory assumes that individuals make rational choices: from a set of alternatives, they choose for the alternative with the highest utility given that the situation meets the assumptions (such as perfect information).

An extensive study by Van Eeten & Bauer (2008) argues that malware is an outcome of the underlying incentive structure in the market. Each actor in the value net makes a rational decision and thereby weights cybersecurity benefits (the minimization of risks, potential loss) against the cost (inconvenience, effort) of taking security measures. Similar to other markets, this results in externalities: home users do not take into account negative effects for other actors in their decision to behave insecurely. A follow-up study focuses on botnets in particular. End users' incentives to clean-up botnet infections are even smaller compared to other malware since the services of compromised devices are often not disrupted (M. van Eeten & Bauer, 2009). A study by Fagan et al. (2016) empirically supports this theory. They conclude that users act rational (since all participants perceived their benefits greater than costs) and that security behavior is more driven by individual concerns than social considerations (causing externalities). Herley (2009) also finds that rejection of security advice is a rational decision and assigns rejection to poor information about the cost/benefit trade-off in security warnings. He, therefore, concludes that users must be more confronted with the actual harms of non-compliance so they can make a more realistic trade-off. Bulgurcu et al. (2010) combine the RCT with the TPB by subdividing each of the three determinants of TPB in an aggregate sum of costs and benefits. A more recent study by Jhaveri et al. (2017) presents a model of the current abuse reporting incentive structure and conclude that voluntary action is at the heart of effective remediation.

Limitations of the RCT also apply for the context of IoT abuse. End users don't act fully rational. First of all, they don't have full information: they are informed about the consequences, but there is uncertainty about the costs of compliance and non-compliance. Secondly, end users have limited cognitive ability: they do not have the time or mental capacity to weigh the two alternatives. Aytes & Connolly (2004) tested this boundedly-rational choice process for risky computer security behavior and concluded that users don't make sensible choices at all and that additional information will thus not improve behavior.

In conclusion: despite the limitations of the RCT, it provides a powerful rationale to understand non-compliance with IoT abuse notifications. Non-compliance is the outcome of externalizations of the actual costs: an end user does not take into account societal costs in her/his trade-off. Therefore, the costs of compliance do not weigh up against the benefit of a clean IoT device. Based on this, one could assume that IoT abuse notifications can be more effective when making an end user aware of the societal benefits of compliance (or cost of non-compliance). However, the study by Fagan et al. (2016) shows that people, although aware, are still likely to neglect societal concerns. Also, Herley (2009) argues that warning should not confront receivers with *worst-case* harm since users must be enabled to make a realistic cost/benefit trade-off based on *actual* harm.

2.4.3 Protection Motivation Theory

The Protection Motivation Theory (PMT) is grounded in fear appeal theories (Rogers, 1975). The theory explains how fear-arousing communication influences behavior through anticipation of a bad outcome and desire for a good outcome. The PMT was often used in health care studies but proved useful in other domains. Initially, Rogers' (1975) theory contained four[3] cognitive processes that 'mediate the effects of the components of fear appeals' which arouse so-called 'protection motivation'. In later studies, a fifth component is added (response costs) and all five processes are structured in two main processes: threat appraisal and coping appraisal. Threat appraisal is the extent to which a person perceives to be threatened, which is determined by:

Perceived severity: 'the size of the potential consequence, should the negative event occur' (Hanus & Wu, 2016, p.4);

Perceived vulnerability: 'the probability of occurrence of a negative event' (Hanus & Wu, 2016, p.4).

Coping appraisal is the extent to which a person believes s/he can cope with the threat given the recommended response. Coping appraisal is determined by:

Response efficacy: 'one's confidence that certain type of behaviors will allow him or her to avoid or minimize the risk of a negative event' (Hanus & Wu, 2016, p.4);

Self-efficacy: 'the degree that s/he believes it is possible to implement the protective behavior' (Vance, Siponen, & Pahnila, 2012, p.190);

Response costs: 'costs to the individual when implementing the protective behavior' (Vance, Siponen, & Pahnila, 2012, p.190).

The model is illustrated in figure 3. The core idea behind protection motivation (hence attitude change) is that one has the desire to minimize the potential harm of perceived threat, and weights that desire against the perceived coping ability (thus a multiplicative relation) (Rogers, 1975). In Rogers' initial theory, attitude change refers to the 'intention to adopt the recommended response' (Rogers, 1975). In later studies by others 'intention' is often ignored, and the five components are assumed to have a direct influence on behavior. Because of its general nature, PMT is applied in many domains among which information security studies. The consulted literature contains 21 articles that apply and test the PMT in the context of (information/computer/etc.) security advice. Most studies have tested all five components of the PMT, often in combination with another theory. The majority of studies show a significant influence of at least three of the PMT components on behavior. However, which of the five components have significant explanatory value vary greatly among these studies.

[3] The initial theory existed of three processes, but this theory was in 1983 revised by Rogers. Similar to the theory of planned behaviour, PMT was then complemented with the self-efficacy theory of Bandura.

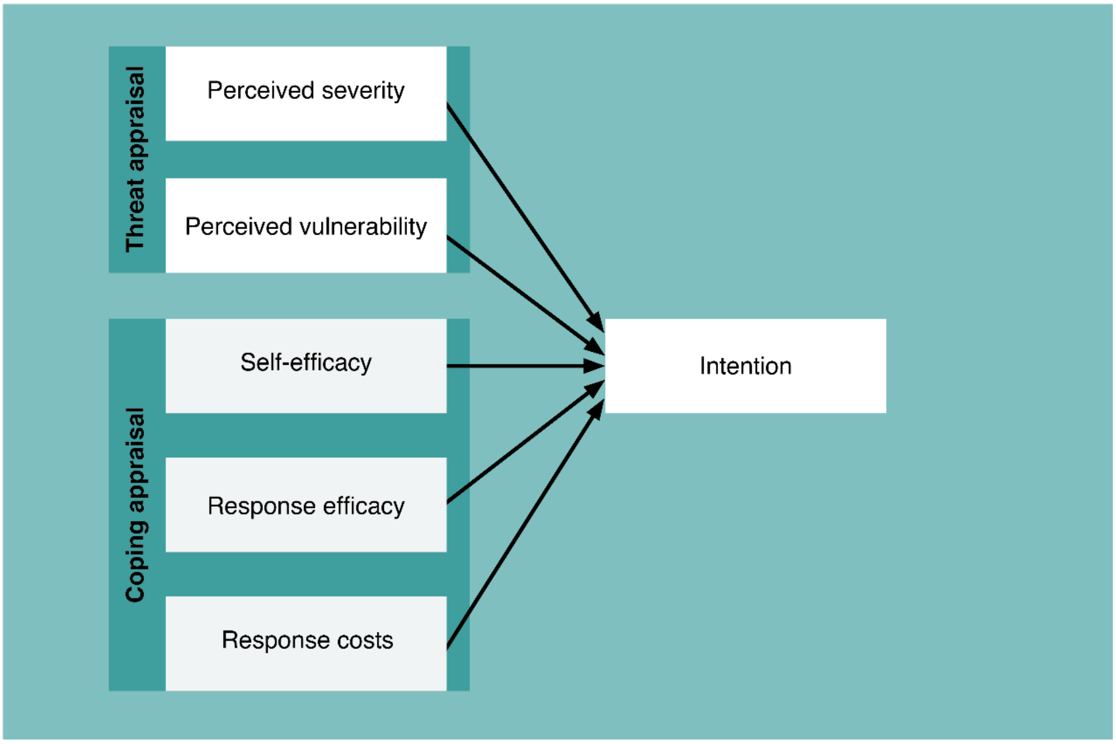


Figure 3 Protection Motivation Theory. Adapted from Rogers (1985)

In conclusion, the PMT provides a model to understand the underlying motivation to comply with a security warning. Key in this model is the notion of *perception,* which implies that not only the information provided in a notification is important, but also its framing. Another relevant assumption in the context of IoT abuse notification is the presence of self-efficacy and response efficacy as predictors of intention. End users must thus be convinced of the efficacy of the recommended remediation guidelines and must also be confident that s/he is capable of performing those actions.

2.4.4 Communication-Human Information Processing model

Warning science (or: risk communication) aims to understand how warnings are processed by a receiver. Models in this domain explain why and when a message is (in)effective and often provide guidelines to design an effective warning. This domain, therefore, provides a helpful perspective to analyze cybersecurity notifications. The Communication-Human Information Processing Model (C-HIP) has been used to this end.

The C-HIP model was introduced by Wogalter (2006) to structure warning research by identifying seven phases between a source that sends a warning and a receiver who will or will not change his/her behavior due to the warning. Wogalter (2006, p.53-58) describes the nine stages the following:

Source: 'the initial transmitter of the warning information'

Channel: the medium and sensory modality (e.g., visual, auditory) in which the warning is sent

Delivery: whether or not the warning has reached its target

Attention switch: the warning must be noticed. It competes with other stimuli from the environment.

Attention maintenance: after notice, attention must be maintained until the message is completely delivered

Comprehension and memory: the receiver must understand the meaning of a message (comprehension) or relevant knowledge must be activated (memory)

Attitudes and beliefs: warning content must concur with what the receiver believes is true (e.g., hazard perception)

Motivation: a warning must energize the receiver to comply

Behavior: whether the receiver carries out the ‘warning-directed safe behavior.’

Environmental stimuli (‘noise’): is not a stage, but captures all aspects other than the warning that may influence how the warning is processed such as other people, other warnings, background noise, etc.

The C-HIP model is illustrated in figure 4. Wogalter (2006) describes the model as a stage or process model, in which information is linearly transferred through each phase. In other words: each phase is a potential bottleneck that could prevent information from being successfully processed. Although a warning may not lead to compliance, it could still have been effective in earlier stages. Although Wogalter emphasizes the linearity of the process, he adds feedback loops to also include the possibility of non-linearity due to processes such as habituation.

This model is applied for the first time for computer security warnings by Egelman et al. (2008). They use the structure of the model in their research design and conclude that active warnings are more effective than passive warnings. Similar to findings discussed in section 2.3.2, Cranor (2008) attribute the failure of compliance with incomplete communication. He argues that receivers are often non-experts who need to be provided with clear instructions. Studies by Felt et al. (2012) and Fagan et al. (2015) focus on the blocking stages of the model. Felt et al. (2012) conclude that permission requests already fail at the beginning stages of the process, namely at ‘attention’ and ‘comprehension’. Fagan et al. (2015) find that update notifications have a negative effect on the ‘attitude/beliefs’ stage due to annoyance, which causes more non-compliance.

Although the C-HIP model is not a widely used model in the cybersecurity domain, the four studies above illustrate that the C-HIP model can be helpful as a framework to understand how the process that leads to (non-)compliance may look like. Due to the structure of sequential stages, it can be an easy tool to ‘pinpoint’ where an end user drops out the process. An interesting notion within the C-HIP model is that notification effectiveness can also be measured based on other stages than compliance only. This way, one could measure the ‘extent of effectiveness’ of notifications rather than the binary distinction between compliance and non-compliance.

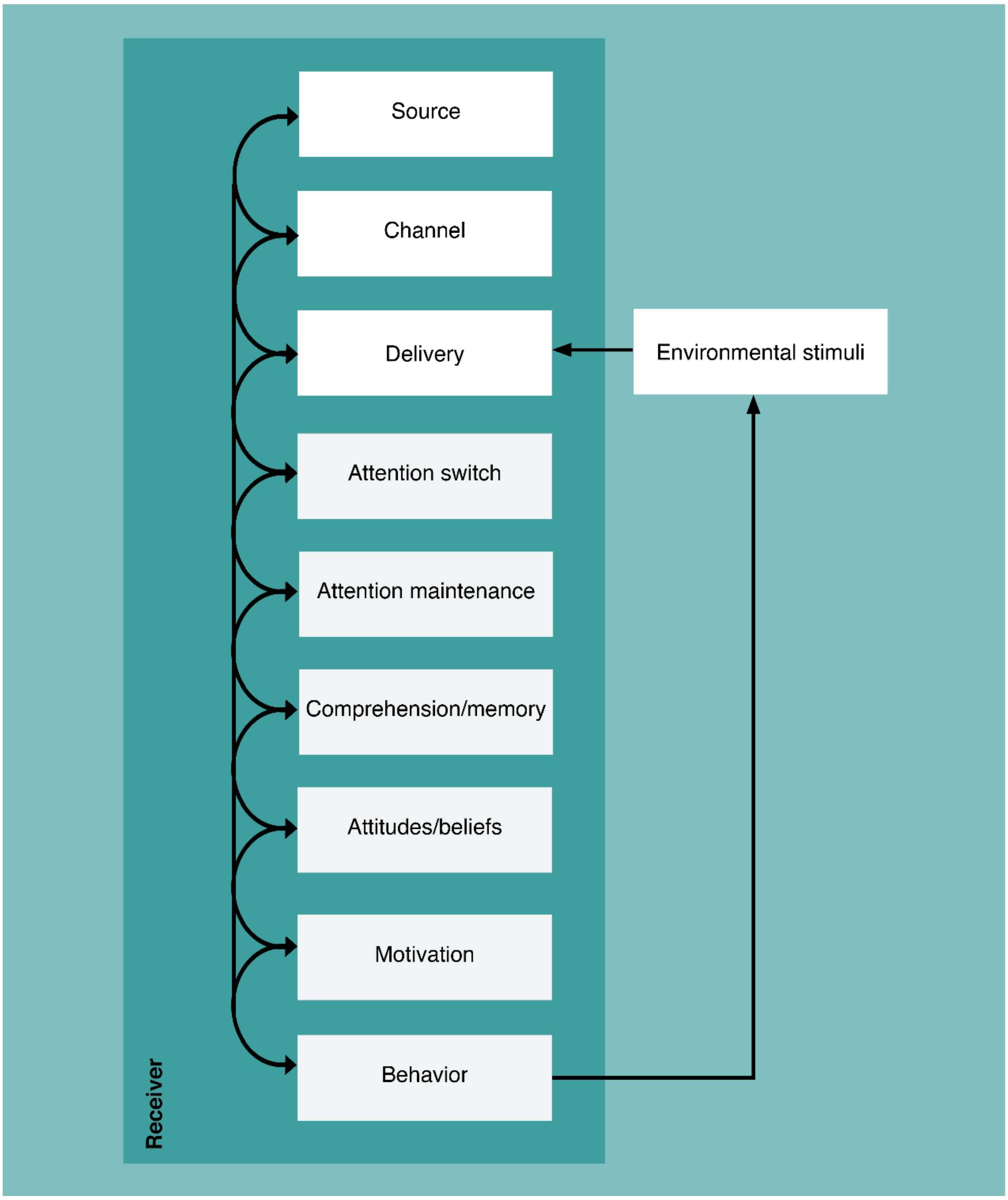


Figure 4 C-HIP model. From Wogalter (2006)

2.4.5 A theoretical framework

The studies discussed in previous sections provide different rationales behind end user compliance with IoT abuse notifications. Whereas the TPM and RCT help to explain end users' incentive to behave safely in general terms, the PMT and C-HIP models acknowledge the influence of the (content of) a notification.

Figure 5 illustrates the combination of these theories adjusted to the context of this study. This model is used to understand the behavior of notified consumers and form the basis for the interviews and subsequent data analysis.

The main structure is derived from the C-HIP model that explains compliance as the outcome of a process with different stages. Attention switch and maintenance are not included as a separate stage but included in the delivery stage since the distinction between the two will be difficult to make within this study. Since 'delivery' now not only refers to the technical aspect of delivery (e.g., an e-mail is successfully sent) but to the fact the message has reached its target (the receiver has *read* the e-mail), we choose to name this stage 'awareness' to avoid possible confusion.

The stage 'motivation' is extended by the PMT since Wogalter (2006, p.58) mentions similar predictors for this stage as the ones in the PMT such as 'cost of compliance' and 'severity of injury'. 'Perceived vulnerability' is not included since the notification addresses an actual infection and is not a vulnerability warning. 'Attitudes and beliefs' are included within the motivation stage. Since motivation and intention are closely linked (someone who is motivated to take action, has also the intention to do so), the motivation stage is not treated as a separate stage but as a further specification of the intention stage.

'Intention' refers to the intention to comply and is derived from TPB and PMT. Both theories argue there is a gap between intentions and behavior. The last stage 'behavior' is subdivided into two stages: 'Compliance' is the behavior-component that refers to whether the end user has actually complied to the recommendations in a notification. Since there are other ways to clean up Mirai than the recommended actions, we add the behavior-component 'cleanup actions' that refers to all other effective cleanup efforts.

Four out of the five variables in TPB are included in this model. 'Subjective norm', which is also recognized as an influencer of the 'motivation' stage in the C-HIP model, is excluded since it addresses social pressure or influence which is not an applicable factor in a home context. Also, the cost-benefit trade-off that is central in the RCT can be derived from the components of the theoretical framework: minimizing the perceived severity versus response costs. To reduce further complexity, the feedback loops between all stages and the loop through the 'environmental stimuli' are excluded.

In conclusion, the main factors which will be used in this research to understand consumer behavior after receiving a notification are:

- Awareness: which entails both the technical delivery of a notification and the awareness of a consumer of the notification content;
- Comprehension: whether a consumer understands what the problem is and what the recommended steps are;
- Intention: whether a consumer intends to comply with the recommended actions in the notification. Intention can be explained by consumers' motivation to comply;
- Behavior: whether a consumer has performed the recommended actions correctly or took other effective cleanup measures.

When a notification has successfully passed all four stages, a Mirai infection is remediated.

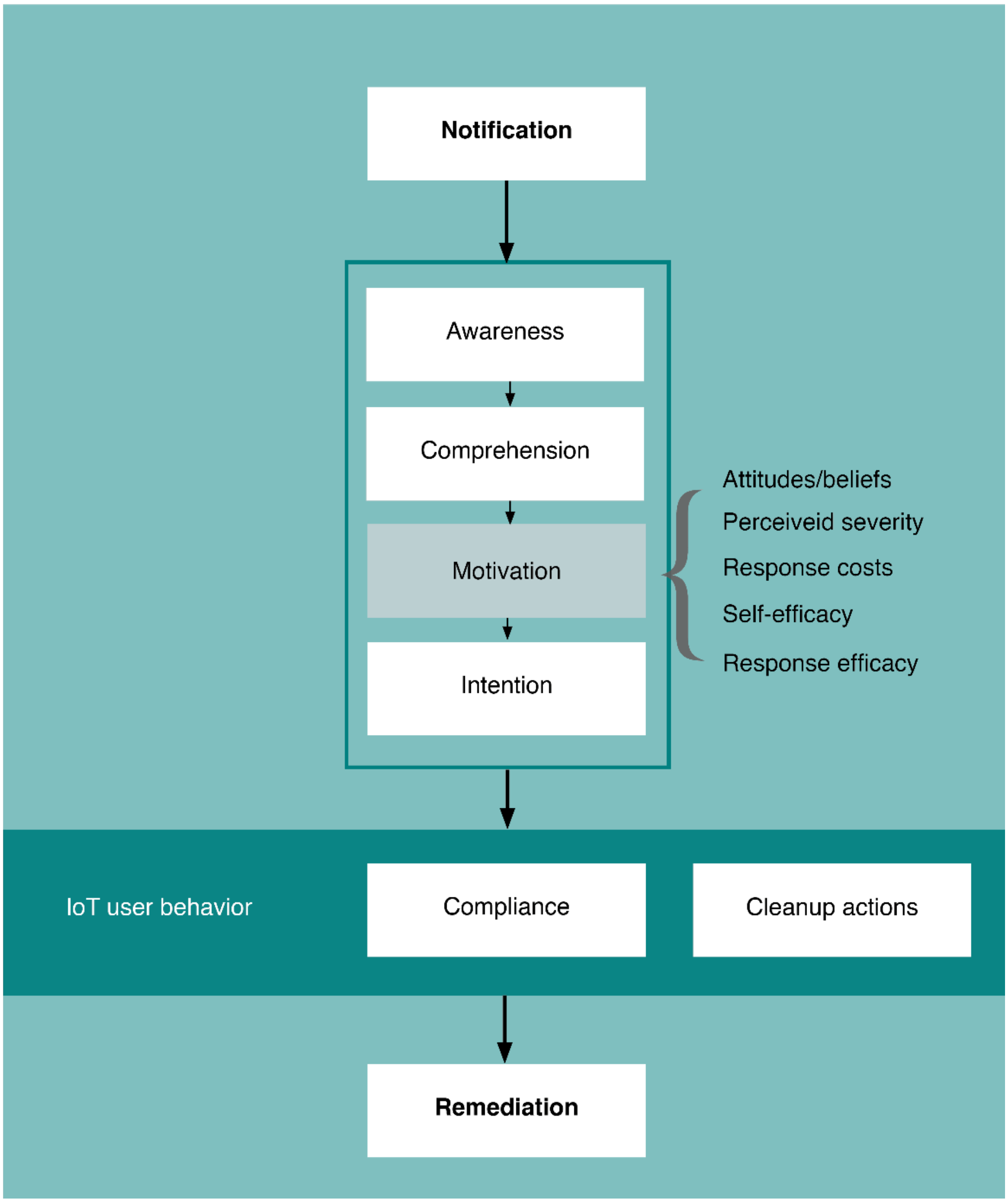


Figure 5 Theoretical framework

3 KPN Abuse Desk

3.1 Introduction

This chapter describes the processes of the KPN Abuse Desk and thereby provides an overview of how infected consumers are notified. The Abuse Desk is an eight-man sub-department of the KPN Security Operation Centre (SOC), which in turn is part of the umbrella department Chief Information Security Office (CISO). The CISO's mission is 'to keep KPN reliable, secure and trusted by customers, partners and society' (internal document CISO, 2019). The Abuse Desk has as a primary goal to remediate vulnerabilities and abuse of KPN resources. This concerns the malicious activities of customers as well as unintentional abuse or vulnerabilities. Abuse Desk employees notify infected customers and mitigate the damage by placing them in a walled garden. Section 3.2 describes how abuse incidents and vulnerabilities are detected. The notification practices are described in section 3.3.

3.2 Detection

The Abuse Desk depends on external organizations for data on abuse incidents, vulnerabilities and other malicious activities. This data is typically provided through so-called abuse feeds. The data providers can be divided into three categories:

- Non-profit organizations which detect and/or collect abuse incidents and notify the concerning ISPs;
- Commercial enterprises that collect and sell abuse data as a service;
- Individuals or individual organizations that 'come across' abuse incident data and report this to KPN.

Of all sources available to KPN, Shadowserver and AbuseHUB provide information about Mirai infections. The Shadowserver Foundation is a non-profit organization that collects a large amount of threat data and sends daily reports to parties such as network providers, governments and law enforcement agencies (Shadowserver, n.d.-a). The KPN Abuse Desk receives every day 41 lists with vulnerabilities and abuse incidents at IP addresses that fall within the range of KPN's and Telfort's networks (Shadowserver, n.d.-b). The 'Drone/Botnet-drone Report' is the only list that contains reports on Mirai infections.

The Abuse Information Exchange is a Dutch Association which represents more than ninety percent of the Dutch ISPs. Their central system, AbuseHUB, collects and correlates data about infections from different sources and share this with the joining ISPs (Abuse Information Exchange, n.d.). In this study, we can only make use of the Shadowserver feed since Mirai detections provided in the AbuseHUB's abuse feeds cannot be easily retrieved. Each abuse incident has to be visually inspected to determine whether the incident concerns Mirai. The Abuse Desk is currently working on a software system that makes it possible to gain overview of the abuse incidents reported by AbuseHUB.

3.3 Notification

The majority of the feeds from Shadowserver are automatically processed. In principle, all reported abuse incidents among KPN consumers are remediated by placing the consumer in a walled garden. This is a quarantined environment from which the consumer can still visit a few whitelisted websites so s/he can perform urgent actions (e.g. financial websites, e-mail hosting websites) and can perform the recommended actions (e.g. website that provides virus scan software). Consumers who are placed in a walled garden receive both an e-mail and landing page in their browser which contain information about the reason for quarantine and recommended steps to remediate the abuse. The e-mail and landing page of Mirai customers are improved using the recommendations by Altena (2018) and are shown in appendix D.3. The recommended steps to remediate Mirai are the following:

1. Identify the devices that are connected to the Internet;
2. Reset the device(s);
3. Change the passwords of the device(s);
4. Reset the modem/router (back to factory settings);
5. Change the password of the modem/router.

Note: this is a concise version of the steps, appendix D.3 shows the complete formulation of the recommended cleanup actions. Consumers in a walled garden can release themselves by filling in a contact form on the landing page. This form is also shown in appendix D.3. This contact form is the same for all abuse incidents and thus contains irrelevant questions for Mirai remediation (e.g., a customer is asked for virus scan logs). Customers can self-release from a walled garden two times. After the second time, consumers have to wait for an employee of the Abuse Desk to release them. When a customer is still in the walled garden after a month, the blockade is automatically lifted.

As previously mentioned, this procedure only concerns KPN consumers. Telfort consumers and KPN business and wholesale customers on KPN's radar are not automatically processed and are notified on a best-effort basis. The procedure is illustrated and described in more details in appendix D.2.

All consumers are asked to e-mail the Abuse Desk when one has performed the steps. In reality, mainly consumers who experience difficulties contact the desk. Consumers cannot call the Abuse Desk, which is a precaution to prevent work overload. The employees can thus only be reached per mail during office hours (Monday to Friday, 8 am – 17 pm).

4 Methodology

4.1 Overview

This research follows an experimental design. It studies the effect of two different notification mechanisms through a randomized controlled experiment setup. This study will be performed for both KPN and Telfort customers, which results in two separate experiments due to possible characteristic differences between these customers.

Sections 4.2 provides information on the experimental setups. The two data collection methods – infection tracking and interviews - are described in respectively section 4.3 and 4.4. These sections will also focus on data preparation procedure. The methods used for data analysis are described in section 4.5. This chapter concludes with an elaboration of the ethical considerations (4.6) and limitations (4.7) of the methodology. Due to the complexity around the estimation of infection time in this study, we visualize this at a conceptual level in figures 8 to 10 at the end of this chapter.

4.2 Experimental setup

In an experimental design, a factor or subject is manipulated to explore its effect. In our experiments, the notification mechanism is the manipulated factor and its effect on customers' behavior is explored. The experiments are *randomized* (random assignment to groups) to minimize selection bias, and thereby increase the validity of the results. The experiments will be *controlled* (inclusion of a control group) to be able to study the causal effect of notifications on consumers' behavior and remediation.

4.2.1 Intervention

Two interventions will be tested: e-mail only notifications and walled gardens. In e-mail only notifications, a customer receives an e-mail which:

- notifies the customer about the Mirai infection;
- provides the five steps to remediate the malware and prevent reinfection as described in section 3.3;
- requests the customer to respond with an e-mail to the Abuse Desk to the notification.

The walled garden notification mechanism consists of the e-mail as mentioned above, but additionally puts the consumer in a quarantined environment. More information about the notification process can be found in chapter 3 and appendix D.

IoT-infected consumers are assigned to either one of the two treatment groups or the control group. The control group will not be notified during the experiment period. When the customer is still detected as Mirai-infected after this period, s/he will receive a notification too.

4.2.2 Populations of interest

The target population in the experiment exists of consumers that own a Mirai-infected IoT device. This research focuses on two consumer markets: KPN and Telfort consumers. Customers from the business, the wholesale and mobile markets are excluded for different reasons:

- Business market: currently, Mirai-infected business customers are notified on best-effort basis. There is no procedure or database in place to match an IP with corresponding business or the right person within that business. Infected IP addresses are randomly selected and attempted to notify (see appendix D). Also, it is not desirable to put a business customer in a walled garden since this may lead to severe economic or safety consequences.
- Wholesale market: in this market, other service providers make use of KPN's infrastructure and network and sell this service under their own brand. The end users of the Internet services thus don't take service directly from KPN. KPN cannot identify nor notify these users directly.
- Mobile market: following our findings from the literature review, Mirai is not a threat to mobile devices (smartphones and tablets). Therefore, this market is not in the scope of this research.

Normally, Mirai infections within the Telfort market are not included in the regular notification procedures and are also remediated on a best effort basis. However, since the identification and notification of Telfort customers are possible, this population will be included in the experiment.

4.2.3 Procedure

On all working days, the Mirai feeds of the previous day are checked for new infections. All Mirai-infected consumers who have been notified before are excluded from the experiment to avoid the influence of habituation. Consumers who are only detected on Fridays or Saturdays are not included because these consumers may not be notified due to the unavailability of the Abuse Desk during the weekend. This is a limitation of the experiment setup and discussed in section 4.7. All other consumers are assigned following two premade lists of complete random assignment. Since KPN and Telfort consumers will be treated as different populations, they are assigned following a separate random assignment process. The procedure for this assignment and corresponding replicable code can be found in appendix F. F Randomization protocol

The experiment subjects are tracked for a period of two weeks. After these two weeks, all consumers will be contacted for an interview by phone. We set this time to two weeks because we want to obtain as reliable information as possible regarding what actions a consumer took. Since the memory of consumers may get blurrier over time, we decided to interview consumers immediately after the experiment period. Section 4.7 on limitations discusses this choice in more detail.

All consumers are tracked for an additional two weeks after the experiment period to monitor whether an IP address is still visible in the abuse feed. If so, that would imply that the Mirai-infection is not remediated during the experiment. Our interviews may have influenced these observations since we may help remember a consumer to cleanup their device, or may alarm a consumer in the control group, which could result in cleanup actions within the two weeks of observation. Also this limitation will be discussed in section 4.7.

During the total of four weeks (two weeks tracking + two weeks of extra observation), a consumer is put on a white-list to prevent him/her from getting a notification about another malware or vulnerability. The experimental procedure is illustrated in figure 6.

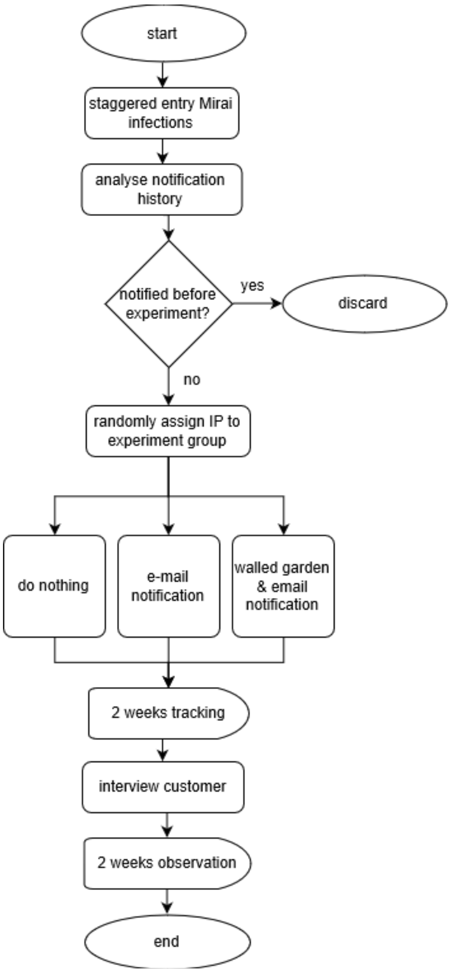

start
staggered entry Mirai infections
analyse notification history
notified before experiment?
yes
discard
no
randomly assign IP to experiment group
do nothing
e-mail notification
walled garden & email notification
2 weeks tracking
interview customer
2 weeks observation
end

Figure 6 Experimental procedure

4.2.4 Experiment duration: an exploration

The observed infections represent the complete population of interest within the experiment period. In other words: we do not take a sample. Due to the absence of sampling variability, inference of the data to a larger population is inapplicable (Neal, 2015). However, this research deals with inferences about the differences between populations due to treatment evaluation (different notification mechanisms).

The experiment has a maximum duration of ten weeks due to time constraints. To determine the minimal number of consumers needed in the experiment to reach significance given sufficient power, we conduct a power analysis. Because of fluctuating numbers of detected Mirai detections per day, it is unclear yet how many consumers in the experiment are to be expected in this period. For that reason, a dynamic power analysis is performed to explore the power level for different population sizes. Since infection time is the most dominant variable, this variable is used for the power analysis.

The power analysis is computed and visualized using the G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007). Since Altena's (2018) study concludes that there is no significant difference between the control group and e-mail notification group, we use the walled garden group to determine the expected effect size. The following input parameters are used for the power analysis (t-test based):

Input parameter	Value	Support
Tail	One	The effect has an expected direction: the walled garden group has lower mean
Effect size	0,4	We cannot obtain the mean values of prior studies and therefore not estimate the effect size. We choose to set the effect size on 0,4, which is a medium effect size.
Alpha	0,05	The probability of wrongfully rejecting the null hypothesis (type I error).
Allocation N2/N1	1	Consumers are complete randomly assigned (thus equal size)

Table 1 Input variables power analysis

The power level is the probability that one does not make a type II error, in this context: the probability that we will not detect a difference between two populations while there actually is. Figure 7 visualizes the relation between population size and power level for this experiment. Although one wishes to maximize the power of its outcomes, there is no consensus on the minimal level. The minimal power level for this research is set to 50%. As can be derived from the figure, the experiment needs more than 105 consumers (70 for two treatment groups, the experiment has 3: 70/2*3). We wish to reach a power level of 80%, which comes down to 234 consumers (156/2*3). In conclusion: to obtain enough data to be able to detect a difference between infection time of two groups given the effect size of 0,4, we need a minimum of 105 Mirai infected consumers and wish to approach 234.

Altena's (2018) experiment contained nine consumers per week average. Her dataset contained only KPN consumers, and all observations were obtained from one source only (Shadowserver). In our experiment, we also include Telfort consumers and use data provided by a second source (Thunderlab, see section 4.3.2). We, therefore, think ten weeks is a long enough experiment duration to reach a minimum of 105 Mirai-infected consumers. If this number is not reached after ten weeks, we have to accept an increased risk of not estimating differences between populations while these exist in reality (type II error).

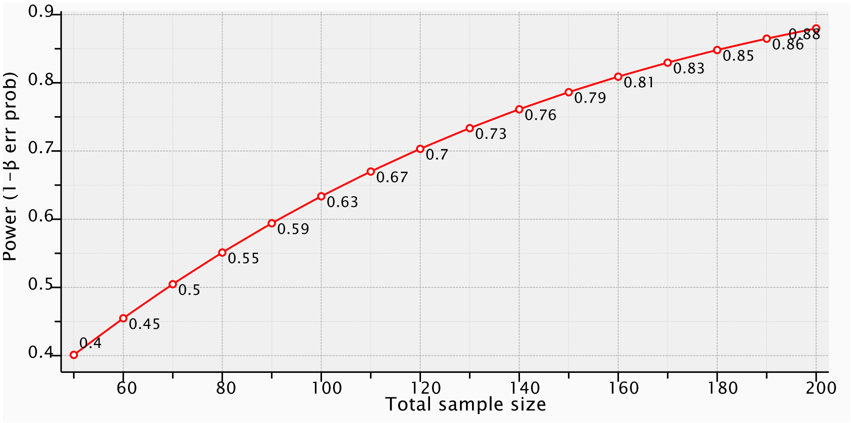


Figure 7 Dynamic power analysis

4.3 Tracking infections

4.3.1 Estimating infection time

The Mirai infection time is estimated using the daily abuse feed of two sources which will be described in section 4.3.2. These feeds provide timestamps of when Mirai is detected at an IP address. In this study, the infection time is estimated by the difference between the time of notification and the last detection. There are exceptions to this rule:

- The last detection of consumers in a walled garden is set to the bailout timestamp since Mirai cannot be detected during this quarantined state. If a consumer has not been able to release him/herself, the timestamp of the first communication with the Abuse Desk is used. If Mirai is detected again after the bailout, this timestamp is denoted as the last detection.
- Since the control group has no notification timestamp, the notification timestamp is set to the first day after detection at 11 am. This is the regular time on which the other treatment groups are notified.
- If Mirai is not detected after notification, the infection time is set to a random number between 1 and 12 hours. We choose to do this - instead of setting the infection time to 0 - since zero-values may lead to assumption violations during Cox modeling (Cox modeling is described in 4.5.2). We choose to randomly pick a number because this more realistic than one specific infection time. Because we want to maintain as much information as possible, we choose for the range 1 to 12 hours and not longer.
- If Mirai is detected in the observation period (after the two-week experiment), the infection time is set to 336 hours (two weeks) and the infection is included in the analysis as a censored observation (censored observations will be explained in section 4.5)

We choose to set the notification timestamp as start time rather than the first detection because notifications can speed up remediation. Also, Altena’s (2018) study uses a similar setup, so replicating this setup permits comparison of the results.

4.3.2 Mirai infection sources

Two sources are at our disposal of this experiment: the Shadowserver Botnet-Drone report and a darknet infrastructure named ‘Thunderlab’.

Shadowserver

The Shadowserver Foundation is a large repository of security information internationally (Shadowserver, n.d.-a) and shares this information freely with network owners. Of all current reports available, Mirai detections are shared in the ‘Botnet-Drone report’. An interview with Rosie Lovell, personas analyst of Shadowsserver, provides more information about their detection methods (see appendix E). Currently, the detection of Mirai is done in four ways:

- An in-house honeypot network of 600 IP addresses;
- A honeypot network of 1000 IP addresses funded by the General Cyber Alliance (GCA);
- 169 sinkholes of Mirai variant 14;
- Third-party feeds (e.g. large ISPs) that provide raw data or fingerprinted data.

Thunderlab

SURF is a Dutch cooperative of educational and research institutions that provides ICT facilities to its community (Surf.nl, n.d.). A network segment of its network SURFnet is made available for research purpose (Surfnet.nl, 2018). This darknet infrastructure is called 'Thunderlab'. This infrastructure provides access to 131,070 IP addresses.

4.4 Interviews

There is a vast body of literature on how to develop and conduct interviews for research purposes. The Interview Protocol Refinement Framework (IPR) by Castillo-Montoya (2016) combines existing resources on conducting research interviews and structures this using a four-phase process. This framework provides a systematic approach to develop interview protocols and thereby increase their reliability.

4.4.1 Phase 1: alignment with research questions

The interview questions need to be aligned with the research questions to obtain 'intentional and necessary' questions (Castillo-Montoya, 2016, p. 812). This alignment can be checked using a matrix that displays what interview question answers which research questions. This matrix shows what questions are unnecessary (not giving an answer to a research question) and if there is a gap (research questions that are not covered).

From the literature review, we have gained an understanding of how and why consumers behave or comply with an abuse notification. The stages of the theoretical framework (section 2.4.5) are used to align the interview questions with.

4.4.2 Phase 2: constructing an inquiry-based conversation

Phase 2 entails the search for balance between conversation and inquiry in an interview. This goal can be reached by making sure the questions meet common interview rules (Castillo-Montoya, 2016):

- The questions are accessible and approachable;
- The interview follows 'social rules that apply to ordinary conversation' (Rubin & Rubin, 2012 p.96 as referred to in Castillo-Montoya, 2016);
- The interview meets its inquiry goals by structuring it by types of questions (introductory/transition/key/closing);
- Likely follow-up questions and prompts are pre-defined in a script.

The questions from phase 1 are adjusted and complemented to meet these guidelines. Appendix G presents the matrix that contains the questions, what topic it covers, the type of question and what the follow-up question will be depending on the answer.

4.4.3 Phase 3: Receiving feedback

Gathering feedback on the interview protocol is key to enhance the reliability and trustworthiness of interviews as a data collection method (Castillo-Montoya, 2016). The interview protocol is reviewed by two researchers of the research department and two employees of the KPN Abuse Desk. One Abuse Desk employee provided little points of improvements. He suggested having the contact details of the Abuse Desk and Help Desk ready for when consumers would like to contact the Abuse Desk in the future and for referral when consumers ask questions regarding their subscription.

4.4.4 Phase 4: Piloting the interview protocol

The last phase of the interview refinement is trying out the research instrument in real life. This pilot tests A) whether the questions lead to the intended answers, B) whether interviewees understand the question, and C) how long an interview takes (Castillo-Montoya, 2016).

KPN consumers that were placed in a walled garden due to Mirai have been interviewed to test the protocol. Of 17 consumers, eight consumers were available for an interview. The following points are taken into account for the final improvement of the protocols:

- Consumers seem to better understand the problem and the recommended actions than their answers to the contact form suggest;
- Consumers overestimate their remediation effort (performed some of the five recommended steps while stating they complied fully);
- Although customers are subscribed to a consumer subscription, some of them use their Internet subscription for business purposes (50%);
- No consumer could recall the brand of their infected device;
- Some consumers (25%) received no landing page and no e-mail and thus were unaware of the notification;
- The interviews took between 5 and 10 minutes each.

4.4.5 Conducting interviews

Prior to each interview, the communication of a consumer with the Abuse Desk and Help Desk is studied. This enables us to conduct better-informed interviews. Due to KPN’s wish to not record interviews with its customers, the interviews cannot be fully transcripted. To capture the data, the answers are written down in a pre-made form during the interview and directly entered in a Python script afterwards. This script automatically asks the correct input based on the treatment of a consumer and his/her previous answers. This makes sure the complete data is entered and cannot be modified accidentally.

When a customer is not reached, a voice-mail is left to inform the consumer that we will attempt to reach the consumer another time. A customer was taken off the interview list after three attempts. We did not communicate the purpose of the call in the voice-mail so consumers who have not been reached, do not know they are Mirai-infected.

4.5 Data analysis

In this study, we are primarily interested in the Mirai infection time, or: the time between ‘birth’ and ‘death’ of a bot. This kind of data is known as ‘time-to-event’ outcomes and includes censored observation. Censored observations arise when a lifespan is longer than the period in which a subject is observed. This is visualized in figure 10 at the end of this chapter. Exclusion of censored observations would lead to information loss when analyzing lifetime probabilities (Klein & Moeschberger, 2003). Because of these conditions, survival analysis is the designated branch of statistics to analyze the data of this study.

Survival analysis is a set of statistical methods wherein the time to an event is the outcome variable and includes censored observations (Klein & Moeschberger, 2003). Three models will be used in this study: the nonparametric Kaplan Meier estimate, the semi-parametric Cox Hazard model and the parametric Accelerated Failure Time (AFT) model. The first two methods are commonly used in survival analysis, the latter is less common but may provide extra insight as will be explained in section 4.5.4.

4.5.1. Kaplan Meier survival curve (nonparametric)

Since malware studies have similar features as clinical trials (treatment groups, infection time), estimating survival probabilities is a common practice to analyze infections (Orçun Çetin et al., 2019). At the base of survival analysis lies the survival function and the hazard function (Klein & Moeschberger, 2003). The survival function provides the probability S that a subject (malware infection) is still alive after time t (remediation X of the bot has not occurred yet).

$$S(t) = \Pr\,(X > t)$$

The Kaplan Meier product limit estimate can provide this survival curve when dealing with censored data (Lindsey et al., 2004). The Kaplan-Meier survival curve is a step-wise function of 'the probability of surviving in a given length of time' (Goel, Khanna, & Kishore, 2010). $\hat{S}(t)$ is the probability a subject still lives before time t, estimated by the number of deaths that happened during the last event d_i and the number of living until that moment n_i. This is formulated as follows:

$$\hat{S}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i})$$

In the context of this study, 'subjects' refers to Mirai bots and 'death' refers to the remediation of Mirai bots. Key in this estimate is the inclusion of partial information: bots that are not remediated after the experiment time of two weeks, are also included in the estimate as shown in figure 10. These cases are referred to as 'right-censored observations' (Goel et al., 2010). The inclusion of right-censored data prevents the underestimation of survival probability.

The Kaplan-Meier estimate has several underlying assumptions of which one is critical to highlight in light of this study. The survival probabilities are assumed to be the same for all infections (Goel et al., 2010). We must thus distinguish curves for all groups of which we assume have different features.

To compare survival behavior, one can compare the survival differences over time (entire curves) or at specific times. Entire curves can be compared using log-rank tests when the assumption of proportional hazards is met (Lifelines, 2019b). This assumption is true when all populations have the same hazard function but with a different ratio (see next section). In case that curves of different population cross, this assumption is thus violated and comparison of the curves does not lead to accurate outcomes. However, time-specific log-rank tests can always be performed (Lifelines, 2019b).

4.5.2 Cox proportional hazard regression (semi-parametric)

The problem of Meier-Kaplan estimates is the exclusion of heterogeneity; in other words: each curve is independently drawn while groups may be dependent. In the context of this study, groups can be related (e.g., similar device types or user characteristics). These shared features can lead to an underestimation of the influence of each variable (O'Quigley, 2018, p.152). In the presence of other covariates which are not taken into account, we may not detect a difference between two survival curves while there actually is. This is the so-called 'omitted variable bias'. We can overcome this by using regression models.

There are several techniques to regress covariates. Cox's proportional hazard regression is a common semi-parametric method in survival analysis and includes covariates (Z). The dependent variable in Cox's model is the hazard rate (HR): the risk of death at the begin of a small time interval, given that a subject has survived until then (Klein & Moeschberger, 2003).

$HR(t_j) = \Pr(X = t_j | X \geq t_j) = \frac{p(t_j)}{S(t_{j-1})}$, j = detections of remediation

The survival function and hazard rate are related by:

$$S(t) = \prod_{t_j \leq t} [1 - HR(t_j)]$$

The Cox proportional hazard model assumes that hazard functions of different groups are proportional to each other: they all have the same baseline hazard function (λ_0) and a partial hazard $\exp\{\beta Z\}$ that is dependent by covariates (O'Quigley, 2018, p. 156). This can be mathematically formulated as:

$$\lambda(t|Z) = \lambda_0(t) \exp\{\beta Z\}$$

Wherein:

$\lambda(t|Z = 0) = \lambda_0$(t)

The parameters (Z) can be estimated by maximizing the partial likelihood of the weights β. By estimating a Cox regression model, we can identify the influence of covariates (e.g., treatment, market) on the survival behavior of Mirai. These covariates (Z) are included in the model as a vector and can take the form of interaction effects and dummy variables. The exponential of the coefficient is the multiplying factor of the hazard function. In other words: a covariate with an estimated coefficient β will have on time t a hazard rate of the mean hazard rate at that time multiplied by exp(β).

A Cox hazard model can only be created when the aforementioned assumption of proportional hazard is met. In some cases, this assumption is violated because the baseline hazard functions of covariates are completely unrelated. Stratification can be applied in these cases so that the baseline hazard function is estimated for each individual stratum and the explanatory value of a covariate can be analyzed (Klein & Moeschberger, 2003, p. 308).

4.5.3 Accelerated Failure Time model (parametric)

The Cox hazard model is most commonly used for survival analysis (Klein & Moeschberger, 2003; Saikia & Barman, 2017). This model needs no specification of a probability distribution. We choose to also include a parametric regression model to analyze survival data, namely the Accelerate failure time (AFT) model which is a popular approach when modeling failure time in a parametric way (Klein & Moeschberger, 2003; Saikia & Barman, 2017). The inclusion of this model A) enables us to be still able to analyze the data when the Cox proportional assumption is violated, and B) provides us with different information than the Cox model.

The AFT model describes the survival function as the product of a fitted baseline survival function and an acceleration function. The acceleration factor determines the change of the time scale of the survival curve compared to the time scale of the baseline and is formulated as $\exp\{\beta Z\}$ (Klein & Moeschberger, 2003, p. 394). This is formulated as:

$$S(x|Z) = S_0[\exp\{\beta Z\}\,\mathrm{t}] \text{ for all } x$$

This model is based on the assumption that the baseline survival function S_0 follows a particular probability distribution. A variety of models can be used to represent S_0. The Weibull distribution is a popular distribution due to its flexibility (hazard rate can be increasing, decreasing or constant) (Klein & Moeschberger, 2003, p. 395). To determine the distribution which must be chosen for the AFT model, the most common distributions are fit on the data. These are exponential, Weibull, Lognormal and Log Logistic distributions (Klein & Moeschberger, 2003).

To determine the best distribution of the baseline survival curve, we use the Akaike Information Criterion (AIC) with P as the number of parameters and K as the number of coefficients (Saikia & Barman, 2017; Zare et al., 2015):

$$AIC = -2(loglikelihood) + (P + K)$$

The model with the lowest AIC-value fits best. Note: when comparing models with an equal amount of parameters, we can directly compare the log-likelihood values of the models.

The estimated acceleration factor is the 'ratio of survival times corresponding to any fixed value of survival time' (Saikia & Barman, 2017, p. 413). An estimated covariate thus 'stretches' or 'shrinks' a survival curve by a constant amount, in other words: a covariate with an estimated coefficient β will have the same survival function on time t as the baseline survival function at time exp(β). Interpretation of this is quite intuitive: the mean and median survival time is multiplied by exp(β).

4.5.4 Explorative modeling

The Cox hazard model and AFT model have different qualities and weaknesses. These are summarised in table 2 (Bradburn, Clark, Love, & Altman, 2003; Klein & Moeschberger, 2003; Saikia & Barman, 2017).

	Cox hazard model	**AFT model**
Main assumption:	Cox's proportional assumption is met	The survival curve is distributed as the specified probability distribution
Interpretation of estimations:	Multiplying factor on hazard rate (time-specific)	Multiplying factor on survival time
Strength:	No need for distribution specification. The baseline hazard curve is based on actual hazard rates (more valid model).	More informative than the Cox model and more efficient (smaller standard errors)
Weakness:	Provides less information than AFT and estimates are less intuitive.	A distribution must be specified for the survival curve. The model is estimated under the assumption that the specified distribution is true.

Table 2 Characteristics of survival regression models

Because of the explanatory nature of this study, and lack of understanding of the influence of covariates, each modeling process starts with the inclusion of all possible covariates. During backward selection, each step a variable is excluded based on either expert opinion (e.g. multicollinearity or low variance) or on the highest p-value (least reliable covariate). Each improvement of the model is tested using the Log-Likelihood ratio test. The last step of each modeling process is the comparison of the best-fit model with a trivial model (a model without covariates). When the Log-Likelihood ratio estimates a significant difference, the model is accepted. The steps of each model cycle are included in the appendices J,K and L.

The Kaplan-Meier curves, Cox regression models and AFT models are developed using the Lifelines (v0.21.3) Python package (“Lifelines,” 2019a). This package also enables us to check the Cox proportional assumption, to include stratification and to fit the AFT baseline survival function accounting for censored observations.

4.6 Ethical considerations

We wish not to violate any ethical or legal norms. Potential objections are discussed in this section.

4.6.1 No treatment

During the experiment, one-third of the consumers with an infected IoT device doesn’t receive any notification (the control group). To mitigate this prejudice, consumers are well-informed about Mirai and how to remediate it during the interview. This is complemented with an e-mail notification when Mirai is still detected after two weeks. This extra effort also applies to other consumers who are not aware of the notification.

4.6.2 Whitelisting

Since customers in the experiment are whitelisted, they will not receive any other notification. To prevent any damage or harm, customers are removed from the whitelist when a severe malware is detected. The severity of an abuse case is assessed by a senior Abuse Desk employee.

4.6.3 Confidentiality

The research is executed in line with the General Data Protection Regulation (GDPR). During the experiment, we use information from the subscription accounts to contact Mirai-infected consumers. Contact details are looked up every time prior to an interview and will thus not be part of the collected data. All other data is stored locally within the KPN network and will not be used for other purposes. The processed data cannot be traced back to individual persons.

4.7 Limitations

4.7.1 Estimation infection time

One major limitation of this experiment is the reliability level of the infection time estimations. This limitation is visualized in figure 9. The underlying problem is how Mirai is detected using honeypot and darknet: a bot is only detected when scanning the IP addresses of these particular infrastructures (colored red in figure 9). This creates three blindfolds:

- The shorter the scanning activity (e.g., due to high DDoS activity), the smaller the chance that a bot will scan an IP of one of our sources and thus stays unnoticed;
- The remediation of a Mirai bot cannot be detected. We can only see incidents of when a bot has been, but are not able to see its ‘birth’ and ‘death’.

- When a Mirai-infected device is switched-off, Mirai is remediated but the device is likely to be reinfected when switched on. Although a device is technically 'clean', we argue that the Mirai has not been successfully remediated.

There is no obvious or easy way to overcome this limitation in a real-life setting: one cannot monitor customer's outgoing Internet traffic due to legal reasons (Article 8 of the European Convention on Human Rights). Also, due to the many Mirai variants and their unpredictable behavior, no studies exist yet that may help us verify how reliable our estimates are.

This limitation is mainly a problem when looking into absolute descriptives such as infection time and remediation rate. The limitation also prevents us from making future predictions. However, when comparing survival regressions, this problem is less of an issue: although we may not perceive the real hazard rates, survival curve and mean infection time, we can assume that the error term of each estimate is similar for each group (because the limitation applies to all subjects) which enables us to isolate the influence of covariates.

The problem of identifying the death of a bot is partly overcome by the two additional weeks of observation (see figure 10). We can conclude that IP addresses that are seen again during this period are not remediated within the two-week tracking and are consequently censored observations.

In addition to detection difficulties, we estimate the start of a Mirai infection using the moment of notification. This choice is made because we want to explore the effect of notifications which must, therefore, be independent from the moment of the first detection. However, this choice could imply that infections that are not detected after a notification, have a wrong estimation of zero hours. The difference with estimations using the first detection has a maximum of 24 hours (since notifications are sent the day after the first detection).

4.7.2 Messy Mirai detection

As discussed in section 2.2, there are many Mirai-like variants. However, we cannot control for the Mirai variant. There are two reasons for this:

- Detection of Mirai is primarily done using packet fingerprinting. The Mirai scanning code includes the characteristic that its TCP packets begin with a sequence that is equal to the IP address of the device it scans. Therefore, other malware that has copied the scanning code of Mirai is wrongfully tagged as Mirai-like.
- Additional information that may have been collected using Honeypots or sinkholes are not shared in the Shadowserver feed.

We thus do not control for Mirai variant and all results are be based on the assumption that all variants have similar behavior.

4.7.3 No enrollment during weekends

In the procedure as described in 4.2.3, we enroll no consumers during the weekend. Since the Abuse Desk is closed on Saturday and Sunday, consumers cannot receive help and therefore placing consumers in a walled garden is irresponsible. However, this would imply that Mirai-infected consumers who are detected for the first time on a Friday or Saturday (note: the feeds have one day delay), would be excluded from the experiment. This exclusion would lead to a bias in the dataset since consumers detected for the first time on a Friday or Saturday may be different than the other Mirai-infected consumers (e.g., are fulltime workers and thus install an IoT device mostly on Friday evenings or Saturdays).

To compromise this bias, these consumers are still enrolled when they are detected again on a different day. That day will be treated as the first day of infection. Since we are not able yet to explain natural remediation among the control group, this solution is not watertight. The estimated infection time is possibly lower because of potential actions consumers perform when not (yet) notified. On the other hand, consumers with a short infection time will not be enrolled since they are not detected anymore after Saturday. In other words: only consumers detected longer than one or two days (depending on the day of the first detection) will be enrolled, and the estimated infection time of these cases will be lower. Despite this inconvenience, believe this is still a better option than excluding these consumers because it results in less bias and larger population size.

4.7.4 Interview bias

When conducting interviews, we can assume that consumers give answers that may diverge from reality. Consumers may have forgotten what actions they performed, may formulate it not precisely or have done things wrongfully (e.g., identified the wrong device). Also, consumers may give answers they believe are desired (to please the interviewer or because they have the feeling of being judged). This limitation is taken into account in the interview protocol design. Most questions are open and in case of vague answers, follow-up questions are asked to obtain more precise information. Also, the conversation is framed as an effort to help the consumer, which may encourage consumers to speak more freely about their difficulties with remediation.

4.7.5 Interviews as a treatment

As explained in section 4.2.3, the interviews may influence the infection time since interviewed consumers may take action in the two weeks after the interview. In the context of the obtained data, that would mean we observe a lower number of censored observations than if we wouldn't have interviewed consumers after two weeks. We assume interviews may have the greatest impact in the control group since most of the consumers in other treatment groups are already alarmed by the notification. Therefore, we will estimate whether there is a significant difference between interviewed and non-interviewed consumers in the control group.

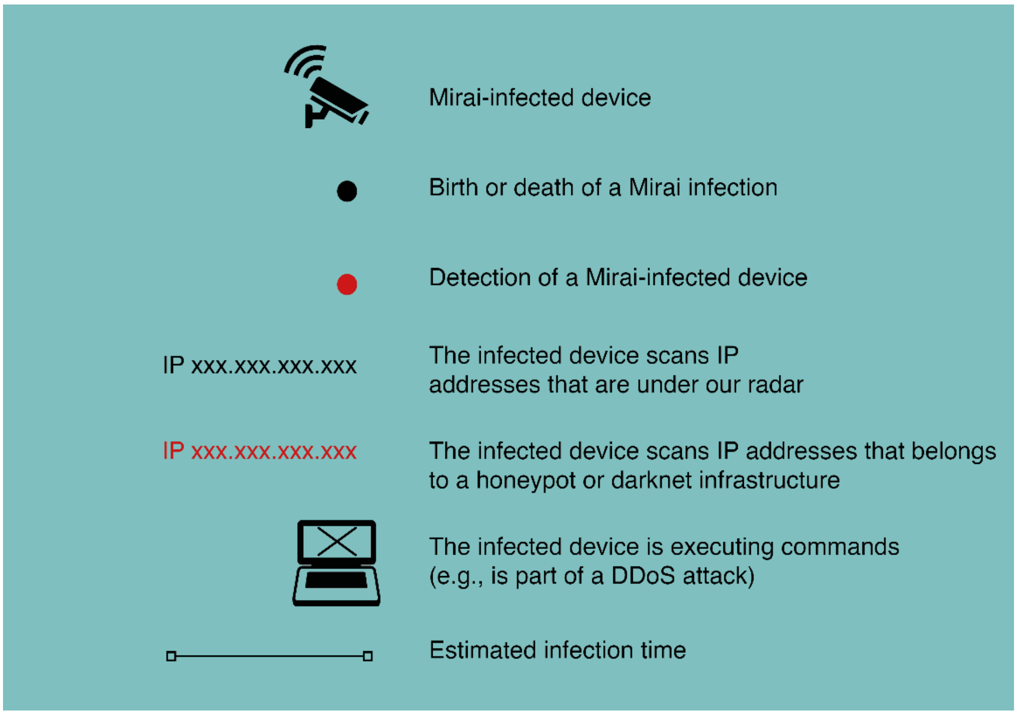

Mirai-infected device
Birth or death of a Mirai infection
Detection of a Mirai-infected device
IP xxx.xxx.xxx.xxx
The infected device scans IP addresses that are under our radar
IP xxx.xxx.xxx.xxx
The infected device scans IP addresses that belongs to a honeypot or darknet infrastructure
The infected device is executing commands (e.g., is part of a DDoS attack)
Estimated infection time

Figure 8 Legend infographics

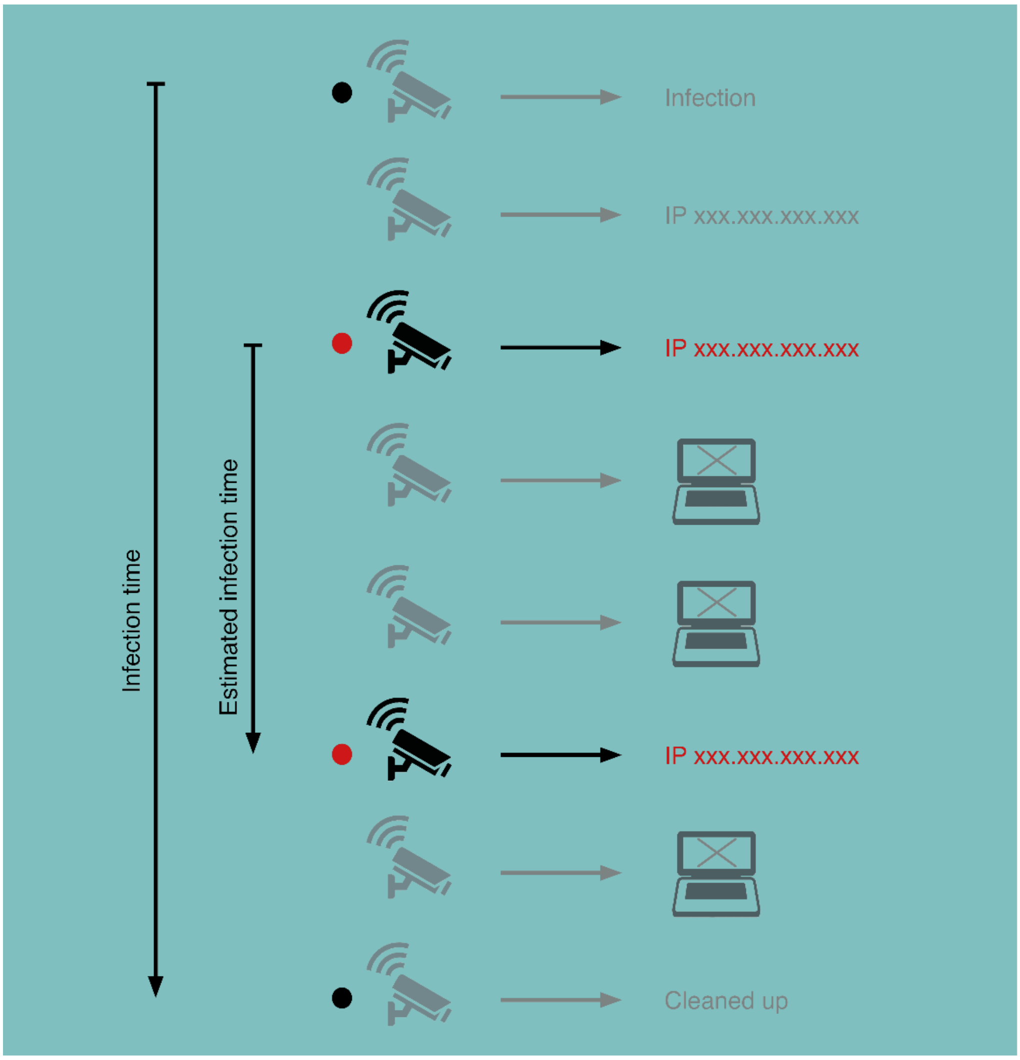


Figure 9 Actual versus estimated infection time

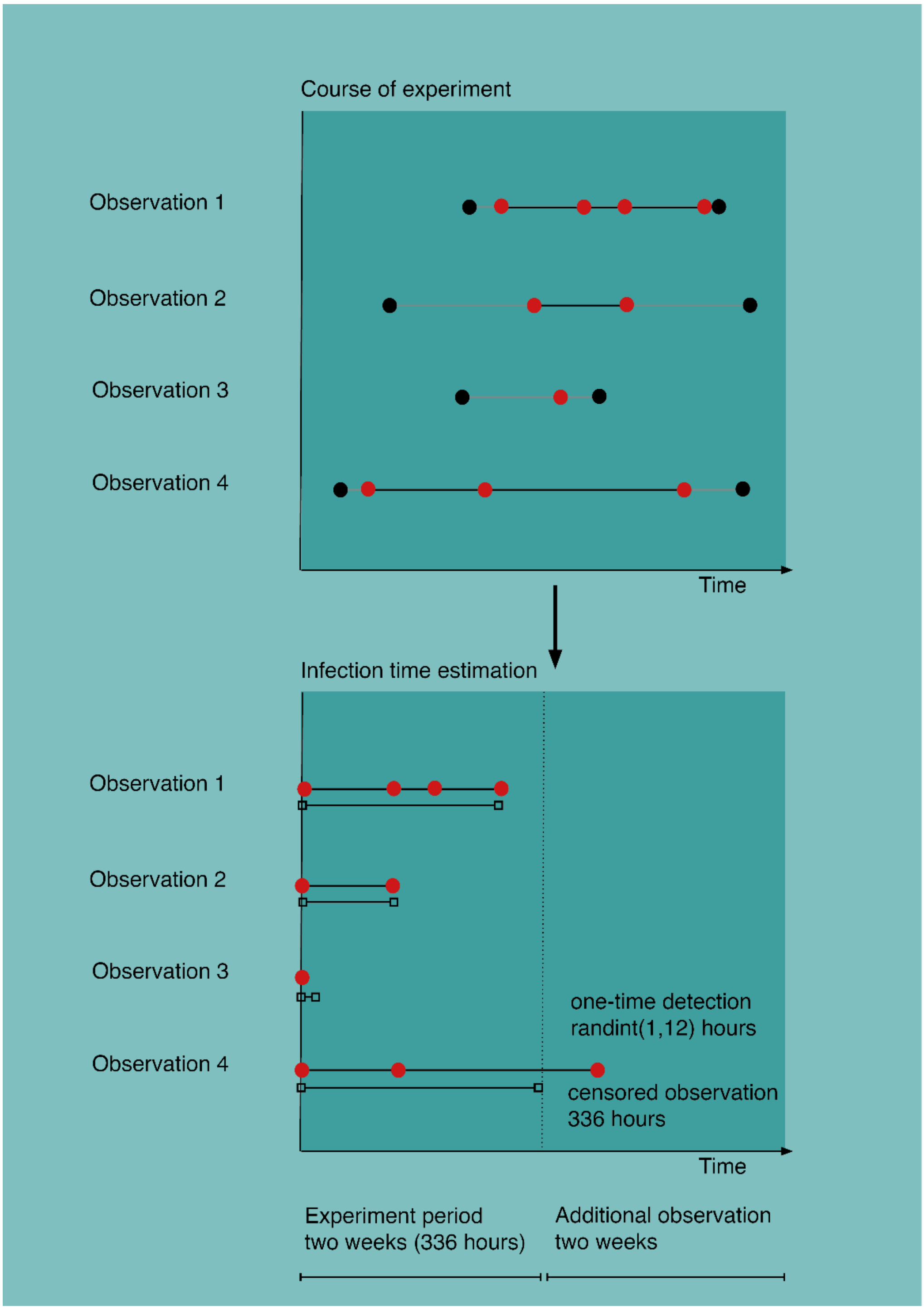


Figure 10 Estimation infection time of normal and censored observations

5 Study population

5.1 Introduction

This chapter describes the enrollment of Mirai-infected consumers in the experiment, the demographics of these consumers, and characteristics of their infected IoT device. This information gives us an overview of the obtained data, and provides us with an answer to the first research question: ‘What are the characteristics of IoT end users who get Mirai-infected?’. Section 5.2 describes the course of the experiment, including an overview of the distribution over the two ISP markets and size of the treatment groups. Section 5.3 describes the demographics of the study population, based on information from subscription accounts. Section 5.4 presents an overview of the identified IoT devices and assignable causes of infection. The data of the latter section is obtained through interviews. Section 5.5 presents the conclusions.

5.2 Course of experiment

The experiment took place from week 19 to week 27, 2019 (eight weeks in total). Newly detected Mirai-infections were included in the experiment within the first six weeks. The last two weeks were used to track the infection time of the latest entries. Figure 11 illustrates the entries of all infections. The green bars illustrate the Mirai infections which are newly detected. The yellow bar represents the infections that have already been detected earlier during the experiment. The numbers of infections are low during the beginning of the experiment, except for the peak on June 1st. From June 10th, there is a peak in the number of new infections which continues the three consecutive days.

We did not receive a Mirai feed from Shadowserver on May 15th and June 2nd. We also didn’t receive feeds from the darknet at the start of the experiment (until May 19th), and on June 16th, 17th, and from June 19th to 24th. The graph shows zero infections between June 20th and June 25th because the Shadowserver feed didn’t contain any Mirai infections (in the consumer markets).

Figure 12 shows the accumulative number of Mirai-infected consumer during the experiment period. Not all Mirai-infected consumers are included in the experiment. Four people were reinfected with Mirai. Five consumers received a notification before the experiment concerning a different abuse incident. One person terminated her KPN contract during the experiment and four consumers were only detected during the weekend. This results in 188 Mirai-infected consumers. During the analysis of the demographics (see section 5.3), eleven consumers appeared to have a business subscription instead of consumer subscription. This leads to 177 Mirai-infected consumers in the experiment.

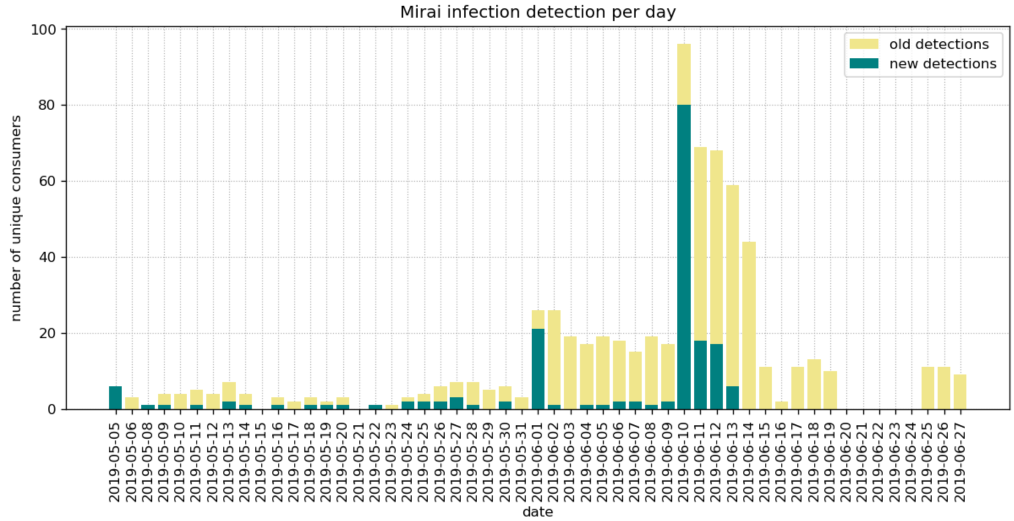


Figure 11 Mirai-infected consumers detected per day

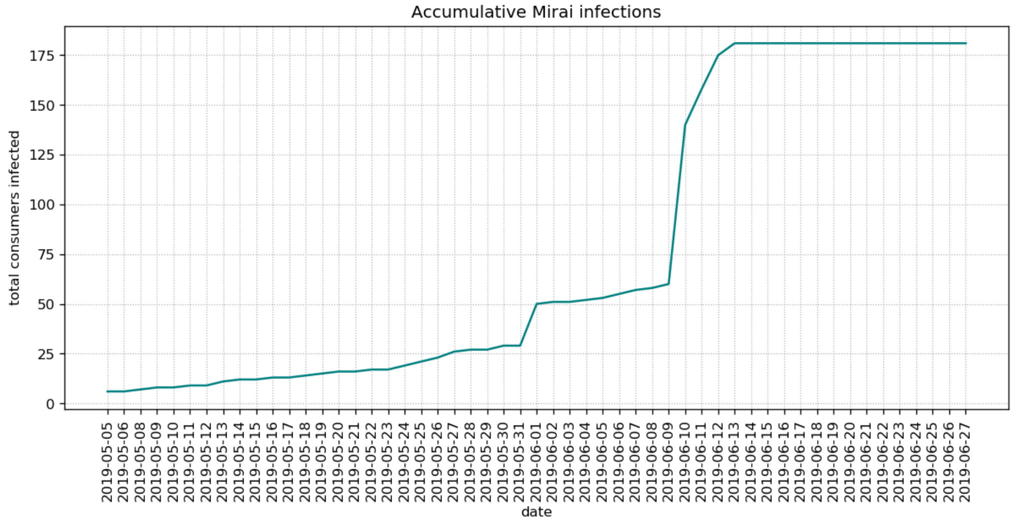


Figure 12 Accumulative Mirai-infected consumers during the experiment

During interviews, it became apparent that none of the consumers of the e-mail notification group had received an e-mail. KPN normally doesn't notify a customer by e-mail only and the mechanism to do this appeared to be malfunctioning since the migration to a new mail server. The consequence of this malfunction is that there is no e-mail group and the control group has doubled in size. The e-mail notification to Telfort consumers did function so that population still has an e-mail treatment group.

In addition, the landing page of the Telfort quarantined environment didn't work properly during the experiment. Many consumers were not able to see the landing page due to a technical problem. The majority of these consumers made a link between the denied Internet access and the received e-mail notification. Therefore, we still treat this population as walled warden group instead of the e-mail group since consumers were incentivized by an Internet disconnect to take action.

Of the total, 72% of the consumers are from the KPN consumer market and 28% from the Telfort consumer market. Of all consumers in the experiment, 57% have been interviewed. The interviewed consumers are quite evenly divided over the different groups (see table 3). One person didn't want to partake in the interview, the rest of the consumers were called three times without success. Table 3 summarizes the distribution of consumers over the two

ISP markets and three treatments. The numbers after the slash refer to the number of consumers interviewed within each group.

	Control	E-mail	Walled garden	Total
KPN	85 / 35 (41%)	-	43 / 28 (65%)	128
Telfort	17 / 10 (59%)	16 / 12 (75%)	16 / 11 (69%)	49
Total	102	16	59	177

Table 3 Consumers in the experiment (/consumers interviewed)

5.3 Demographics of Mirai-infected subscribers

The subscription accounts contain information on the demographics of the Internet subscribers. We obtained the gender and birth year for each consumer. Of the KPN consumers in the experiment, 69% are male, 15% female and 7% had a shared account. Eleven consumers (9%) have a business subscription and are therefore excluded from further analysis. The distinction between the different markets is not accurate for some IP ranges, which explains these eleven cases. When looking at the gender distribution of Mirai-infected Telfort consumers, we find that the majority is male (88%), followed by female subscribers (10%). The gender of 2% is unknown.

Figure 13 compares these percentages with the distribution of all Internet subscribers in both markets. The share of male subscribers is in both markets higher than in the overall population. This difference is largest in the Telfort market (30%), and a bit smaller in the KPN market (10%). Using the N-1 Chi-squared test, we find a significant difference between the share of male subscribers among Mirai-infected consumers and the total population (p=0,0002, CI:15-35% in Telfort market, p=0,019, CI:2-20% in KPN market). We don’t estimate a significant difference between the shares of female subscribers.

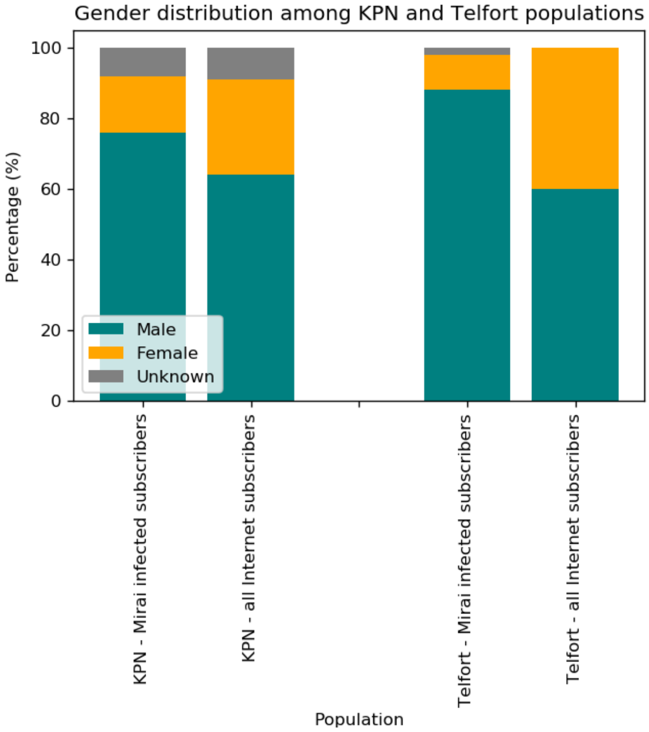


Figure 13 Comparison of gender distributions

The distribution of Mirai-infected consumers over birth years is visualized using boxplots and histograms in figure 14 in the left column. This distribution is wide-spread (between 1932 and 1993), which means that Mirai-infected consumers have greatly varying ages. Ages within the Telfort population are a bit more densely distributed but still show great variety. The median birth year for KPN and Telfort is respectively 1971 and 1972, and all Mirai-infected consumers are older than 25 years.

When comparing these distributions with the birth year distributions of all Internet subscribers (right column of figure 14), we observe that the Mirai-infected KPN consumers have a positively shifted distribution compared with the distribution of all KPN subscribers. The group of Mirai-infected consumers is thus relatively young. The mean age of Mirai-infected consumers is seven years younger (mean age of 48) than the mean age in the total population of KPN Internet subscribers (mean age of 55). Welch's unequal variance t-test estimates a significant difference ($p<0,001$)[4]. This shift is not present in the distribution of Mirai-infected Telfort consumers. There is no difference between the mean age of infected Telfort subscribers and the total population (both 49 years). However, relatively more consumers between 1965 and 1985 are infected.

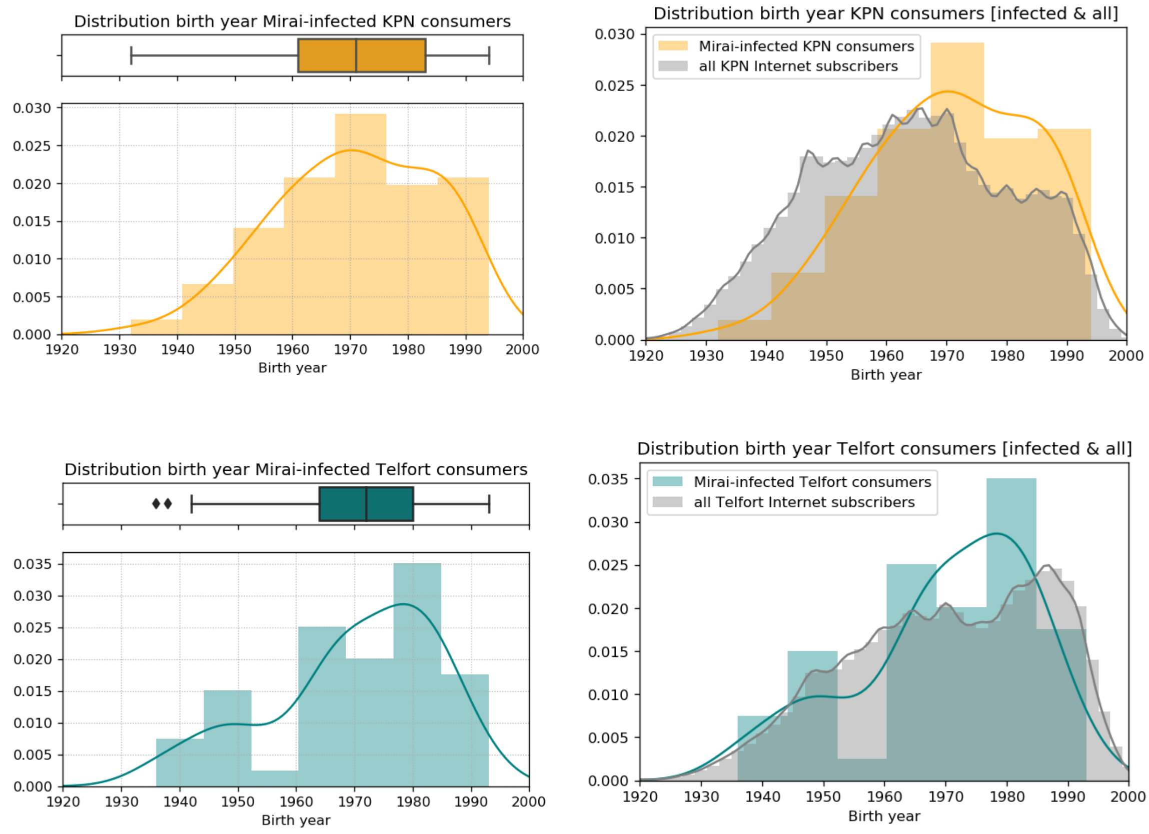


Figure 14 Distribution of birth year per consumer market (left: Mirai-infected consumers, right: comparison with all Internet subscribers)

The subscriber of an Internet service does not per se have to be the owner of the infected IoT device. However, none of the 99 interviewed consumers indicated that the device in question

[4] The Welch's t-test is performed under the assumption that the influence of the Mirai-infected consumers is negligible because of the small number compared with the total population. This population can therefore be regarded as the non-infected population, resulting in two independent groups (infected and non-infected consumers).

was owned or used by another user of the home network. Therefore, the demographic data we obtain from the subscription accounts are a reliable representation of the demographics of the infected IoT device users.

5.4 Identified devices and cause of infection

Using the information obtained from the interviews, we can review the device types that Mirai-infected consumers own and whether they can identify an assignable cause for the infection. Figure 15 displays the percentages of how often a device type is mentioned as possibly infected. The majority (72%) of consumers own an IP camera or Raspberry Pi, followed by NAS (7%) and DVR (6%) devices. 5% of the interviewed consumers could not recall having any IoT devices. The number of Raspberry Pi devices is striking since the recent study by Cetin et al. (2019) identified no Rasberry Pi among 88 infected devices. Also, the identification of a heat pump has never been reported before as far as we're aware.

The high number of Mirai-infected Rasberry Pi devices can be explained by a Domoticz software vulnerability. Normally, Mirai infects IoT devices through brute-force attacks using default or common credentials. Mirai-infections can thus not be attributed to a vulnerability. However, this is different in case of infected devices running on Domoticz software. The Mirai variant on these devices exploits the 'Unauthenticated Remote Command Execution' vulnerability so that it can bypass authentication (Carretto, 2019). This vulnerability was already detected at the end of April 2019 in Domoticz software older than version 4.10577 (Exploit Database, 2019). A new version without the vulnerability is released on the 9th of May ("Download," 2019). Domoticz software runs on home automation devices, often Raspberry Pi and NAS devices. This 'Domoticz-variant' of Mirai explains the high peak on June 10th and the following days.

Since we don't have information about the Mirai-variant of the infected devices, we can only study the outbreak of the Domoticz-variant using the date of the outbreak. Table 4 shows the number of infected consumers per market before and during the outbreak. The period after June 9th (during the outbreak) accounts for two-thirds of the number of infections in the experiment. The number of Telfort consumers that is detected even quadrupled after the Domoticz-variant outbreak.

	Before June 9th	After June 9th	Total
KPN	49	83	132
Telfort	9	40	49
Total	58	123	

Table 4 Consumers per market before and after June 9th

Figure 16 presents the proportion of assignable causes. 45% of the interviewed consumer knew they had a device running on outdated Domoticz software. Except for this group, most consumers could not point out an assignable cause for the infection (42%). 6% of the interviewees had installed a new device. 3% had installed a new Experiabox (a KPN router+modem), 2% connected their device to the Internet, and 2% reinstalled a device that had been temporarily not in use. The installation of a new Experiabox is not an obvious cause for infection. We provide two possible explanations: A) A consumer needs to reconfigure his/her network and does so less secure than before (e.g. using a demilitarized zone (DMZ) or Universal Plug and Play (UPnP)). B) A consumer is assigned to a new IP address and now falls within the observed IP ranges (a few KPN IP ranges are not well categorized). Employees of the Abuse Desk explain that his latter situation occurs rarely.

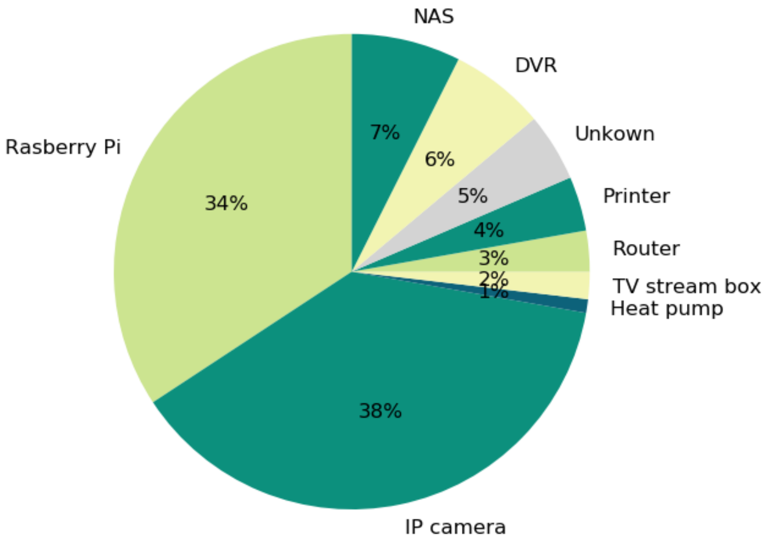


Figure 15 Device types

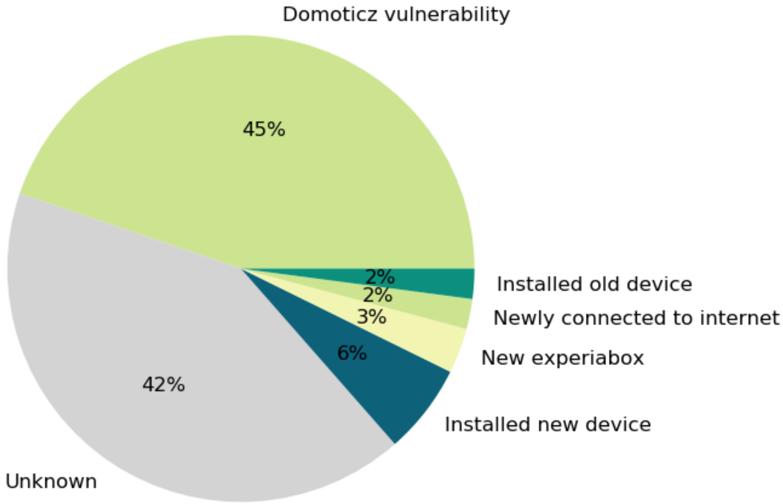


Figure 16 Assignable causes for Mirai infections

5.5 Sub-conclusions on the study population

This chapter describes the study population by looking into the enrollment of the Mirai infections, the demographics of the device users, the identified IoT devices and assignable causes of infection. The goal was to answer the research question: 'What are the characteristics of IoT device end users who get Mirai-infected?’. The data used to answers this question is obtained through subscription accounts and interviews. As addressed in section 5.3, we can assume that the demographics of the subscribers are a reliable representation of the demographics of the Mirai-infected device users.

The population of Mirai infected consumers contains relatively many male consumers when compared with the total population of Internet subscribers. Men are thus more exposed to IoT abuse for which we provide two possible explanations:

- More IoT devices per capita: men are more often in possession of an IoT device compared to women which increase the chance that the owner of an infected device is male;

- More Mirai infections per device: men use their IoT device differently than women (e.g., they use it for more technically advanced applications, or use a device in a less secure manner), which increases the chance on abuse among male consumers.

The birth years of Mirai-infected consumers vary from 1932 to 1993, and the distribution of is widely spread. This is remarkable in two ways: the eldest Mirai-infected consumer is 87, which is quite a high age for someone using something novel as an IoT device. On the other side: the youngest consumer with a Mirai-infected device is 26 years old, which is older than one might expect for the youngest infected IoT users. These findings can be explained when comparing the distribution of Mirai-infected consumers with the distribution of the complete population of Internet subscribers. We observe a great overlap between those distributions. In terms of spread, this implies that the population of infected consumers is a moderately good representation of the complete population. We observe two major deviations:

We conclude that Mirai-infected KPN consumers are relatively younger. The mean age of this group is seven years younger than the mean age among all KPN Internet subscribers. Consumers within the ages of 29 and 54 years are typically more infected than consumers of other ages. Consumers older than 54 are typically less infected.

Telfort consumers within between the ages of 34 and 54 are relatively more infected than others. In contrast to the consumers in the KPN market, there is no difference in the mean age between infected and non-infected Telfort consumers.

During the experiment, we encountered an outbreak of a specific Mirai variant targeting software that runs on outdated Domoticz software. Different than conventional Mirai, this variant doesn’t access the device through a brute-force attack but rather bypasses authentication by exploiting a vulnerability on the outdated versions of Domoticz. This outbreak explains the high amount of Rasberry Pi devices among the infected devices. Telfort consumers have relatively fallen more victim of this variant, which means these consumers are more exposed to Mirai (i.e., are more often in possession of devices running on Domoticz software or using it differently than KPN consumers).

42% of the consumers could not assign the infection to a cause, which is strikingly high because we have no reason to believe that these devices are left out before in scanning activities of other bots. We provide two theories which can explain the unobserved causes:

- The information provided by the consumer is incorrect. A consumer can have misidentified the device or forgot events that explain the cause of infection;
- Recent Mirai-variants use new sets of credentials for their brute-force attack or exploit new/other vulnerabilities which explains why a device is suddenly ‘exposed’ to Mirai. The outbreak of the Domoticz-variant supports this theory.

All in all, these findings provide a general understanding of the context of the problem and are the first exploratory step in identifying patterns that may be of interest for further analysis.

6 Tracking results

6.1 Introduction

This chapter reports on the results of the experiment so we can analyze the survival behavior of different populations. The goal of this analysis is two-fold: we want to obtain an understanding of the data before analyzing it in-depth in the next chapter, and we want to check whether natural remediation - as found in Altena's (2018) study - is also observed in our data (since this is one of the primary motivations for this study). Section 6.2 describes the distribution of infection time per treatment group and ISP market. Section 6.3 further analyses these results by providing remediation rates and survival curves. Section 6.4 explores the effect of the Domoticz variant outbreak on survival behavior. Section 6.5 presents the conclusions.

6.2 Infection time per treatment and market

Before performing survival analysis, we take a look at the data distributions. These plots visualize the center and spread of the distribution of infection time structured by the two markets (figure 17) and three treatment groups (figure 18).

KPN and Telfort consumers show similarities in the distribution of infection time in figure 17: the majority of Mirai bots survive less than four days with a peak of remediation within the first day. Quite a substantial share of consumers is still infected after the experiment period, resulting in the peak at 336 hours (these are censored observations). The low number of remediation after four days is remarkable since it shows that Mirai is typically remediated either within four days or not at all. The boxplots illustrate that the infection times of KPN consumers is wider spread than of Telfort consumers. Due to the higher density at the beginning of the Telfort distribution, the censored observations are considered outliers. Since these are of importance in survival analysis, these will not be excluded.

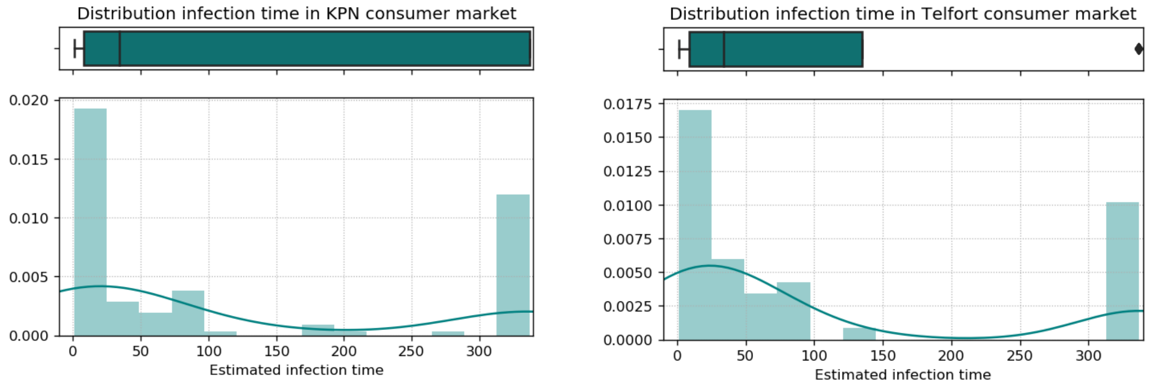


Figure 17 Distribution of infection time per market

Figure 18 displays three plots that show the spread of the distribution of infection time per treatment group. Please note that the e-mail treatment group only exist of Telfort consumers due to the failed KPN notification mechanism.

The control group has the widest spread in infection time and a very high peak at 336 hours, which means that the control group has many censored observations. The e-mail group also has a wide-spread distribution but has a lower median than the control group. The observation in the walled garden groups is very dense compared to the other treatment groups. The number of censored observation is low, thus most of the Mirai-infections are remediated during the experiment. As observed in figure 17, remediation rarely occurs after four days. In figure 18 we see that these cases of remediation can be assigned to observations in the control group.

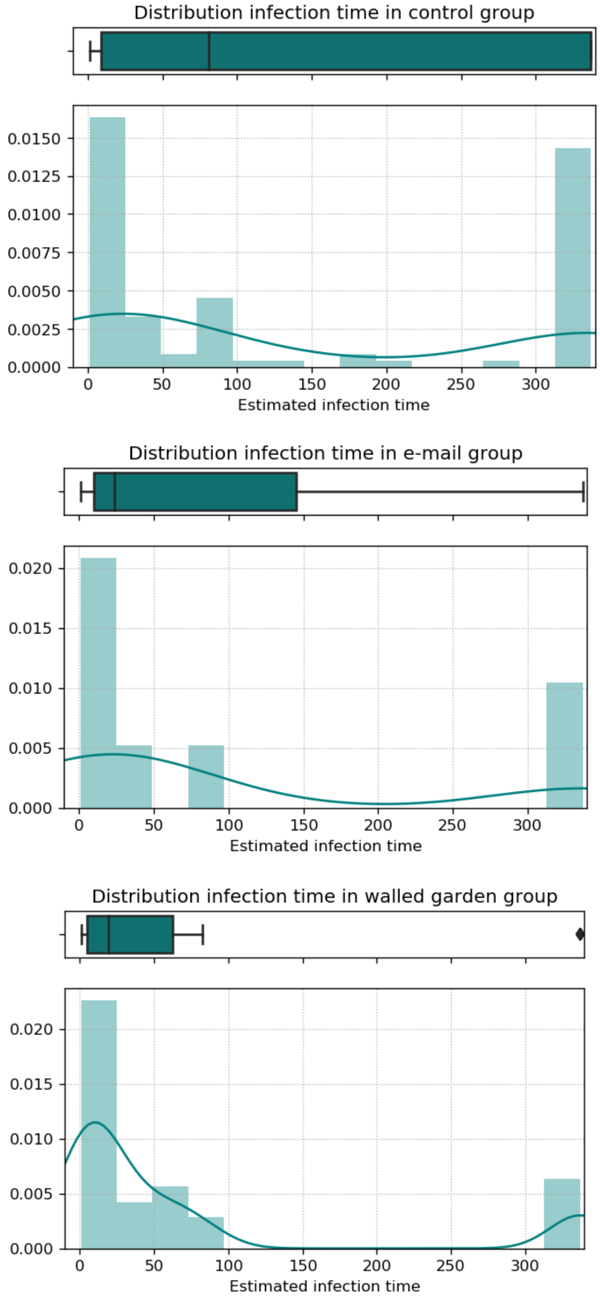


Figure 18 Distribution of infection time per treatment group

6.3 Remediation speed and rate

The density plots in section 6.2 provide information on the spread of infection time per treatment and ISP market separately. This information combined is presented in graph 19 using survival curves. The numbers between brackets in the legend refer to the population size. Please note that the first 12 hours of the survival curve is manipulated: we changed zero-values into a random number between 1 and 12 hours. Section 4.3.1 of the methodology describes this choice.

Figure 19 shows that Mirai infections in the walled garden groups have a higher remediation rate than the other treatment groups. The two survival curves of the KPN consumer market are most divergent; the curves of the Telfort consumer market differ less and are more moderate when compared with the KPN curves: the remediation rate is higher in the control group and lower in the walled garden group. The survival curve of the e-mail group (Telfort only) lies in the middle of the two other treatment groups.

After further visual inspection of the survival curves in figure 19, one can notice a drop after 81 hours. This is reflected in all curves, in some more than others. This drop implies that the chance of survival decreases greatly from one moment to the other, regardless of the market or treatment.

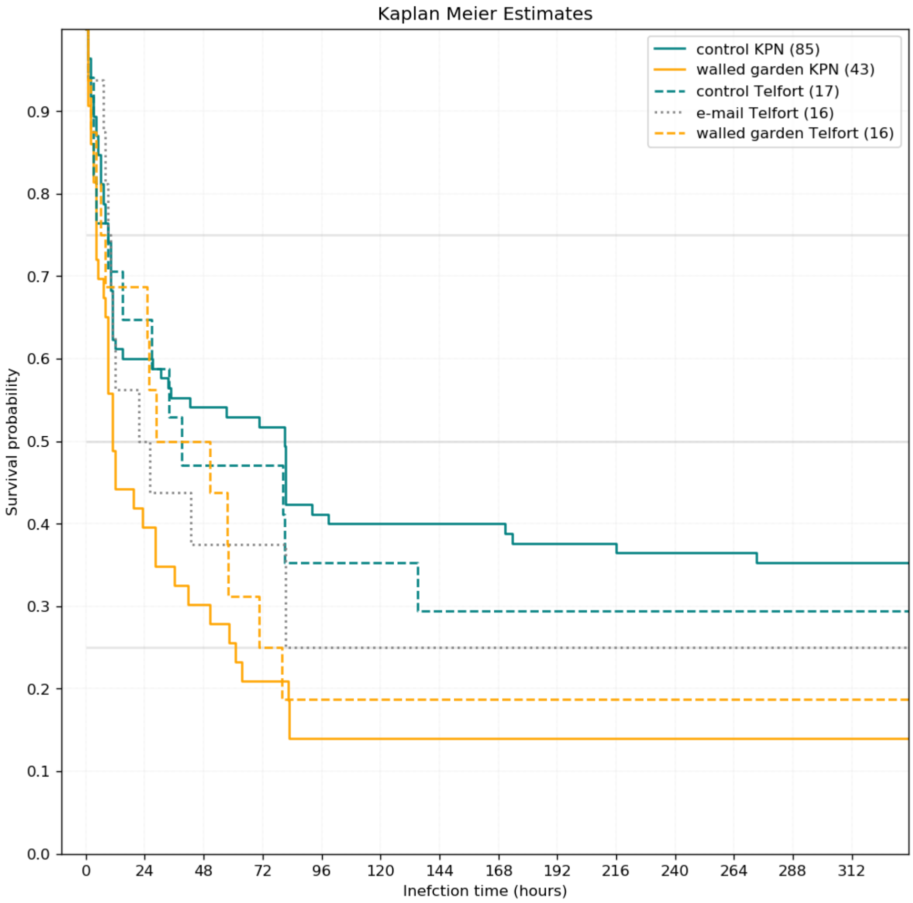


Figure 19 Kaplan Meier survival curves per population and treatment

Tables 5 and 6 contain the remediation rates at three specific times (1, 5 and 14 days) and measures of central tendency. Within the KPN market, the walled garden group has a higher remediation rate than the control group on all three moments. The median infection time is shorter than 12 hours and three days shorter than the median infection time of the control group. The proportional hazard assumption is met for the treatment variable which allows us to do a log-rank test that compares the complete curves. The difference between the two curves is highly significant (p<0,005). The grey values in table 5 are the results obtained from Altena's (2018) study. Almost all remediation rates from that study are higher than the rates found in this experiment. The high remediation rate of the walled garden (97% remediation after two weeks) is not reached in this study (11% lower). However, when comparing the proportions using a chi-square test, none of these differences are significant (see p-values in grey). Striking is the relatively low median infection time in Altena's (2018) study, which is half the median infection time of this experiment's KPN control group (40 versus 81 hours).

The remediation rates and speed of the Telfort consumer market in table 6 show a remarkably low remediation rate in the walled garden group after one day (31%) compared to the other treatment groups and the walled garden group of KPN consumers. Also, the median infection time (40 hours) is higher than that of the other groups. On the longer term, the walled garden group performs better in terms of remediation: more consumers have remediated after two weeks. The turning points are at one and half day (with the control group) and two and a half day (with the e-mail group). The proportional hazard assumption is met for both curves, so we can estimate log-rank tests for the three curves. All the differences between the three curves are insignificant.

	#customers	>1 day	>5 days	>14 days	median	Mean + (std.error)
Control	85	40 %	60 %	65 %	81	145 (149)
Control from Altena (2018)	33	46 % p=0,55	58 % p=0,84	79 % p=0,14	40	-
Walled garden	43	60 %	86 %	86 %	11	66 (112)
Walled garden from Altena (2018)	30	60 % p=1,0	90 % p=0,61	97 % p=0,12	17	-
Log-rank test control vs walled garden survival curve	Proportional hazard assumption is met. Log-rank test estimates a t-value of 9.18 (p<0,005).					

Table 5 KPN remediation rates

	#customers	>1 day	>5 days	>14 days	median	Mean + (std.error)
Control	17	35%	65 %	71 %	39	124 (141)
E-mail notification	16	50%	75 %	75 %	24	104 (137)
Walled garden	16	31 %	82 %	81 %	40	89 (121)
Log-rank test control vs e-mail survival curve	Proportional hazard assumption is met. Log-rank test estimates a t-value of 0.11 (p=0,74).					
Log-rank test control vs walled garden survival curve	Proportional hazard assumption is met. Log-rank test estimates a t-value of 0.58 (p=0,45).					
Log-rank test e-mail vs walled garden survival curve	Proportional hazard assumption is met. Log-rank test estimates a t-value of 0,09 (p=0,76).					

Table 6 Telfort remediation rates

6.4 Exploration Domoticz-variant outbreak

The previous section shows a drop in all survival curves around 81 hours. When combining this observation with the information provided in chapter 5 about the outbreak of the Domoticz-variant, we can isolate the cause of the drop. Figure 20 shows the survival curves of Mirai bots before the outbreak (June 9th) and during (after June 9th). As described in chapter 5, we separate the data based on the date of the first detection because we cannot distinguish different Mirai variants.

The two curves show that the drop around 81 hours can be assigned to the Domoticz-variant. This curve (orange) shows a decrease in survival probability of 12%. In other words: 12% of the bots are last detected after 81 hours. This sudden decrease in survival probability is too big to be coincidental. When exploring the individual observations, it appears that the last detection of these and four other observations (19 in total) is on June 14th between 4 pm and 7 pm UTC. Eleven other observations that are censored also show a disruption from that moment. When adding up these observations, a total of 30 bots (25%) are under the radar (completely or temporarily) at the same time, indicating that they simultaneously stopped their scanning activity.

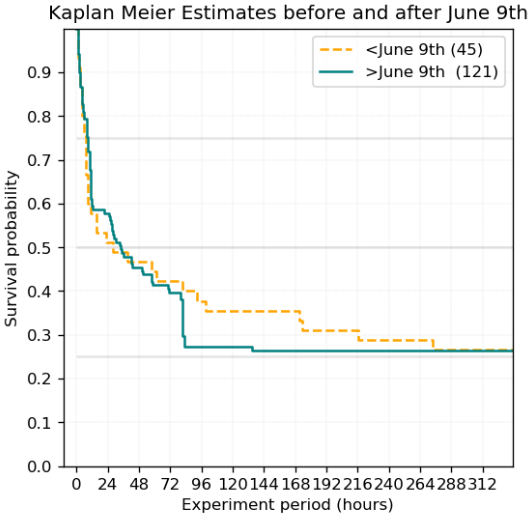


Figure 20 Kaplan Meier survival curves before and after June 9th

6.5 Sub-conclusions on tracking results

The goals of this chapter are to obtain an understanding of the data and to compare the remediation behavior with findings of the previous experiment by Altena (2018).

The results of our experiment are in line with Altena's (2018) findings. Although remediation among our experiment subjects is lower, the remediation rates do not significantly differ with that of Altena's (2018) results. In addition, we find a significant difference between remediation within the KPN control group and KPN walled garden group and a substantial natural remediation rate within the control group (more than 65% is remediated after two weeks).

The infected devices that are remediated have an infection time of typically less than four days. The majority of Mirai-infections are thus remediated within four days, or not at all.

The remediation rate among consumers in the walled group is higher than the other treatment groups. One exception is the first forty hours after notification in the Telfort market: the remediation rate is lower for the walled garden. In hindsight, we know that the Telfort landing page malfunctioned. This obstructed self-release and may have confused consumers, leading to a longer period before the consumer knows s/he must contact the Abuse Desk for the release from the walled garden.

The control and walled garden group of the KPN population are most divergent and significantly differ. Remediation among Telfort consumers is less influenced by notifications than among KPN consumers. Although the survival probabilities of the Telfort treatment groups are divergent, their difference is insignificant. However, we cannot conclude that notifications in the Telfort market are ineffective. The presented survival curves only take the treatment into account, while other variables may have explanatory value. If that is the case, the effect of these omitted variables is attributed to the treatment only, leading to a bias in our estimates (known as the so-called 'omitted variables bias'). To obtain reliable results on the effect of the treatments, we make use of modeling techniques that include more variables. This is presented in chapter 10.

From the analysis of the Domoticz-variant outbreak, we can conclude that a quarter of the bots during this outbreak were given a command by the same botnet herder on June 14th. Nineteen of these infected devices (63%) are not detected anymore after this day, which implies that they are cleaned up but we do not know when (the moment of ceased scanning activity is not the moment of cleanup). The other 37% is detected again after the experiment period and thus are censored observations. It is unlikely that this event is a major obstruction for further analysis because A) it concerns 15% of all infected devices, and B) due to the inclusion of censored observation, we still know how many infections are remediated between June 14th and the end of the experiment.

The occurrence of this event on June 14th gives us an interesting insight into the influence of the attacker on bot behavior. Firstly, it seems that not all bots in the botnet of the Domoticz-variant are deployed for an activity at the same time (since we only identify 30 bots that have ceased scanning activity). Secondly, the censored observations have a remarkably long period of non-scanning (at least ten days). This suggests that either the bots are deployed for long-term activities (e.g. crypto mining), or that they are put on-hold between the execution of short activities.

7 Cleanup efforts

7.1 Introduction

In this chapter, we focus on consumers' behavior. We want to understand what consumers do after receiving a notification, and what unnotified consumers do that may explain the observed natural remediation. The central research question in this chapter is: 'What actions do Mirai-infected consumers perform?'. The answer to this can both help to understand how people remediate, what may cause remediation in the control group, and what difficulties consumers perceive.

7.2 Unnotified consumers

In total, 45 consumers within the control group are interviewed. None of these consumers were aware that they owned an IoT device infected with Mirai. However, there were a number of consumers who experienced troubles with their device and/or Wifi-connection. Four of these consumers contacted the Help Desk. The Help Desk didn't make a link with a possible malware infection and helped the consumers differently: one consumer was told to update his DVR (which was effective), one consumer received a new Experiabox (which was effective only for a few hours), one consumer was sent a KPN technician who inactivated three of the four surveillance cameras (which was effective) and one consumer was not helped at all. Two consumers experienced problems with their Raspberry Pi and decided to reinstall the newest version. One of these two consumers asked during the interview whether 'Mirai could also have caused the bad WiFi connection he experienced' around that same time.

In addition to these two consumers who experienced problems with their Raspberry Pi, six other consumers with either a Rasberry Pi or NAS running on Domoticz software updated their device between the first detection and the moment of the interview. For them, this was just normal routine, not motivated by perceived troubles. Three of them explained that their device is automatically updated whenever a new version is released. However, this statement is questionable since a successfully updated device after the release would not have been vulnerable to infection (see section 5.4: the patched version was released on May 9th).

Strikingly, 41 of the 45 interviewed consumers in the control group were able to identify IoT devices in their home during the phone call. Only three consumers couldn't name any IoT device in their possession and could not recall any visitors who brought any device temporarily.

The majority of consumers in the control group (71%), did not recall doing anything with their IoT device which could explain remediation. This contradicts with the findings of Alterna (2018) and the findings in this study (see chapter 6): we found that 65% of the KPN consumers and 71% of Telfort consumers in the control group had remediated Mirai after two weeks. The remediation rate of the interviewed consumers in the control groups is only 62% after two weeks. We now know that we cannot explain this remediation entirely by the actions of the control group. Only ten of the 45 consumers took correct measures, which explains a remediation rate of 22%. This still leaves a gap of 40% in remediation we cannot explain.

However, one must note that the interviews took place after the experiment period, but still during the observation period. This could have alarmed consumers and thereby stimulated them to take action within the observation period. In other words: the interview is a form of treatment which could have led to less censored observations and thereby a higher remediation rate (see section 4.7 for research limitations). Figure 21 illustrates the survival curves of the interviewed and not interviewed consumers in the control group. The non-interviewed consumers were sent an e-mail only when they were detected again during the observation period (and thus received only treatment when they were already censored). The interviewed group has a slightly lower remediation rate and speed. There is no significant difference between the two groups (Log-rank test estimates a p-value of 0.37). This invalidates the assumption that the interviews have a significant influence on remediation.

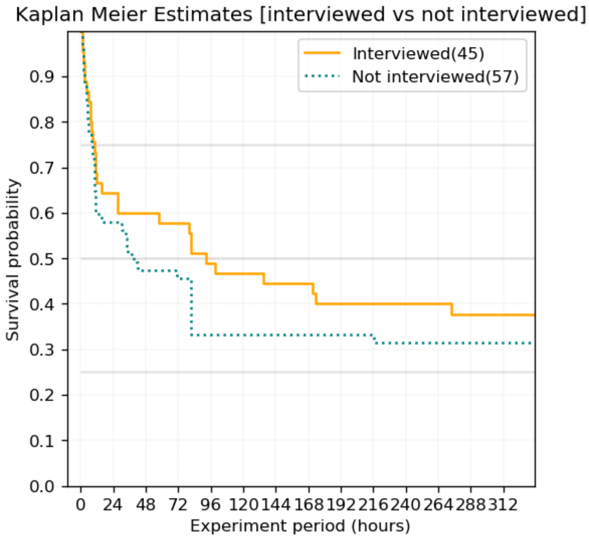


Figure 21 Survival curves for interviewed versus not interviewed control group

7.3 Consumers who received an e-mail

Twelve consumers of the e-mail group were interviewed of whom two did not see the notification since it was sent to an old e-mail address. One of these two consumers cleaned up his infected NAS because 'it worked very slowly'. Of the eight consumers who were aware of the e-mail, five e-mailed the Abuse Desk back for additional questions. All questions asked in the e-mails show a basic technical understanding of the problem ('Do you have the Mac address of the device in question', 'is it possible that this infection is in my Raspberry Pi?) and some e-mails all expressed the wish to keep their devices clean ('Can you warn me again if it happens again?', 'Can I scan for Mirai myself?'). One consumer in particular stands out for his commitment to clean up the infected device. This consumer possesses thirty IoT devices and e-mailed te Abuse Desk regularly to give an update of his cleanup activities. He reinstalled all his devices, performed several virus scans (which is not effective for Mirai), disabled forwarding, and then individually disconnected each device to find the infected devices. The DVR appeared to be the source of the problem.

Two consumers (17%) contacted the Help Desk regarding the notification but were not helped adequately. Both consumers were explained where to download a virus scan and how to clean their computer. One of these consumers knew these actions were incorrect and decided to disconnect his DVR. Another consumer contacted the Help Desk even before receiving the notification because of malfunctioning WiFi. There is only one consumer that did receive the e-mail but did not do anything; he intended to comply but forgot to do so.

The actions of consumers in the e-mail group vary greatly. Only a minority follows all steps that are recommended (16%). The majority performed only some of the recommended measures (22%) or other measures than were mentioned in the notification (22%); or both (33%). Of the recommended measures, a reset of the modem is the least performed action (41%). Some consumers indicate that they don't want to lose their configurations and therefore try to clean up their device without a modem reset. However, it is likely that in many cases, the configuration of the modem is part of the problem. Several e-mailed consumers took more rigorous measures such as disconnecting the device from the Internet or discarding the device completely.

7.4 Consumers placed in a walled garden

Thirty-nine consumers of the walled garden group are interviewed. More than a third (41%) of these consumers called the Help Desk. Nine of these consumers called the Help Desk immediately after noticing the walled garden to ask what to do ('I seem to have a Mirai virus and want to be in contact with the Abuse Team'), three consumers performed the actions and wondered when their Internet connection would be restored, and two consumers were not aware of the notification and asked if there was an Internet outage at KPN. Only one consumer inquired about the authenticity of the notification; she was afraid it might be a phishing e-mail.

All consumers sent a reaction to the Abuse Desk except for one. This one particular consumer was not aware of the fact that she was placed in quarantine: 'we haven't used the computer for months, but indeed, the surveillance camera is malfunctioning already for quite some time'. Only eight consumers managed to release themselves from the walled garden using the contact form. The rest sent an e-mail to the Abuse Desk.

Remarkably many consumers placed in the walled garden choose to disconnect their device from the Internet or to not use the device at all any more, instead of following the steps. Among these consumers, some do this to give themselves some time to take the actions ('my wive has disconnected all devices, I will change the passwords when I'm home'), some because they already doubted the security of the device ('I disconnected the Chinese IP camera and brought it straight to the recycling dump').

Several consumers have difficulties performing the actions. The troubles vary between identifying the right device ('There are no other laptops connected, what do I forget?'), to executing the actions ('How can I know how to change the passwords on my printer?') to fear for the consequences ('I use a lot of home automation but I'm not capable myself to open the ports again after a reset'). Similar to the communication within the e-mail group, many consumers show commitment to remediate Mirai. However, many of these consumers are driven by the disconnect from the Internet rather than concern. Some consumers explain why they need the Internet connection back. One consumer has a security system that doesn't function without Internet access. Another consumer didn't understand what happened so his son took care of the issue. The son expressed his concern and dissatisfaction about the Internet disconnect ('Can you imagine what happens if my parents need the emergency button!?').

23% of the interviewed consumers in the walled garden group followed all recommendations of the notification. The majority of some of these steps in combination with other actions such as a Domoticz software update (17%) and disabling of port forwarding (10%). Consumers in the walled garden can release themselves by filling in a standard contact form which also asks for a virus scan log. This is contradicting the recommended steps and also not an applicable remediation action. Despite this misleading contact form, only 10% of the interviewed consumers indicate specifically to have performed a virus scan. Half of those did that in combination with other actions. Only two consumers performed a virus scan without having seen the contact form and thus made this cleanup effort on their own initiative. However, both consumers seemed to be aware that a virus scan is not effective for Mirai ('It is not possible to install a virus scanner on my webcam'), but did it just to be sure in case it would be effective.

7.5 Performed actions

Figure 22 presents the actions that interviewed consumers have performed to remediate Mirai per treatment group. The first five actions are the recommendations that are provided in the notifications. When looking at the percentages of these bars, we see that these steps are only performed by around half (40-70%) of consumers who received a notification. The consumers in the walled garden do not seem to perform more recommended actions than those who only received an e-mail. Although consumers in the control group were able to identify the IoT devices they own, these are not included in the graph since they didn't identify the infected device *during* the experiment as being Mirai-infected.

Several consumers decide to take other or additional measures: they disable port forwarding (cannot use the device from outside their home network), they disconnect the device completely from the Internet or decide to not use the infected device anymore. Relatively many consumers placed in walled garden decide to perform these 'other' steps than recommended. Two consumers explicitly tell that they have brought the infected device to a recycling dump.

The last bar contains the percentage of consumers who did not perform any remediation step. This is lowest for the walled warden group (2%), followed by the mail group (8%). Within the control group, 71% of the consumers did make any cleanup efforts.

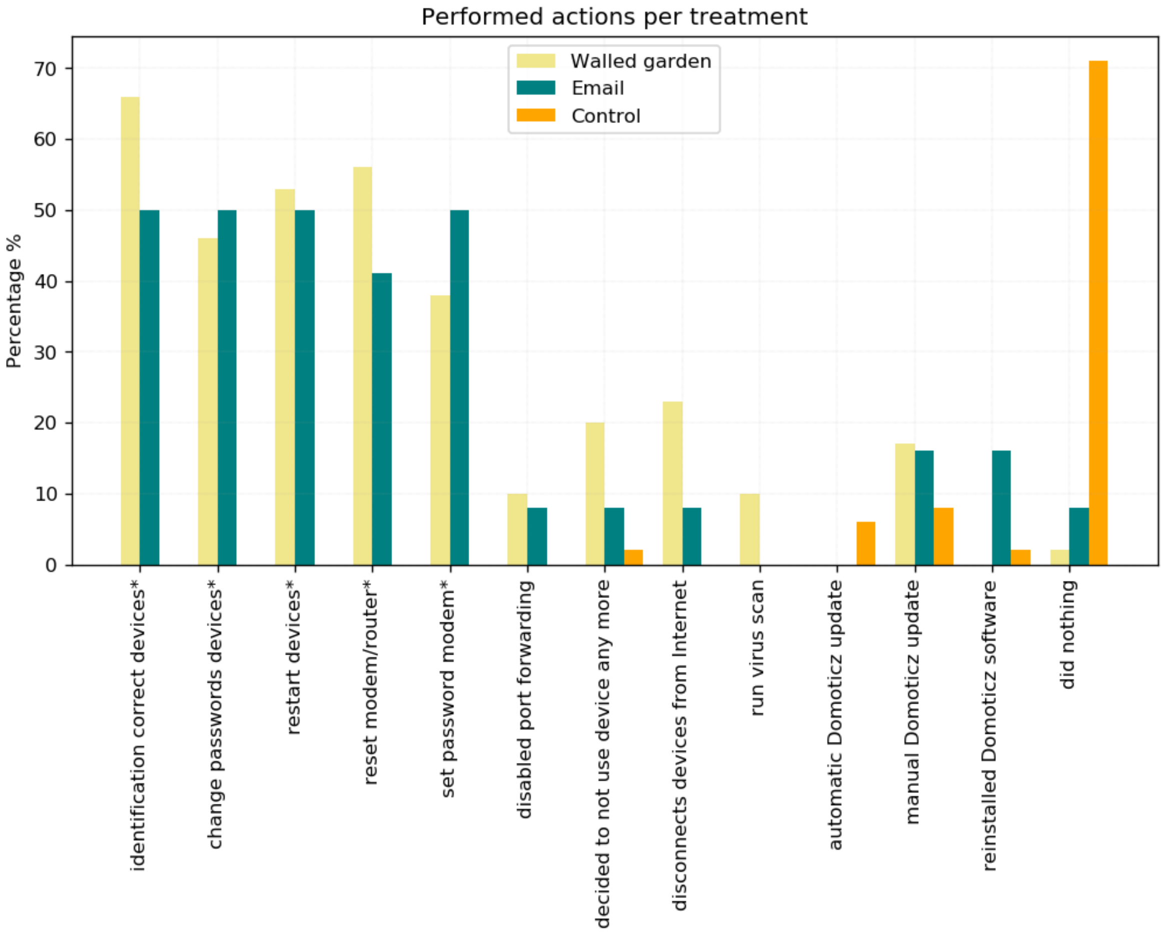


Figure 22 Performed actions per treatment (* is a recommended action in the notification)

Figure 22 provides information per action but does not account for combinations of remediation steps. Figure 23 provides more insight by structuring the remediation efforts in three categories:

- All: consumers who performed all five recommended steps. This category can be described as the consumers who have strictly complied with the notification;
- Some: consumers who performed some of the five recommended steps;
- Other: consumers who have performed other steps than were recommended in the notification.

The left side of figure 23 shows that the majority of consumers who received an e-mail (and did something) has performed some and other steps (33%). Only 22% have complied to all steps in the notification. 11% of all consumers only perform steps other than recommended. The walled garden group (right-hand side of figure 23) also has a majority that executes some of the recommended steps in combination with other actions (40%). A quarter complies completely with the notification and more than three-quarters perform other actions.

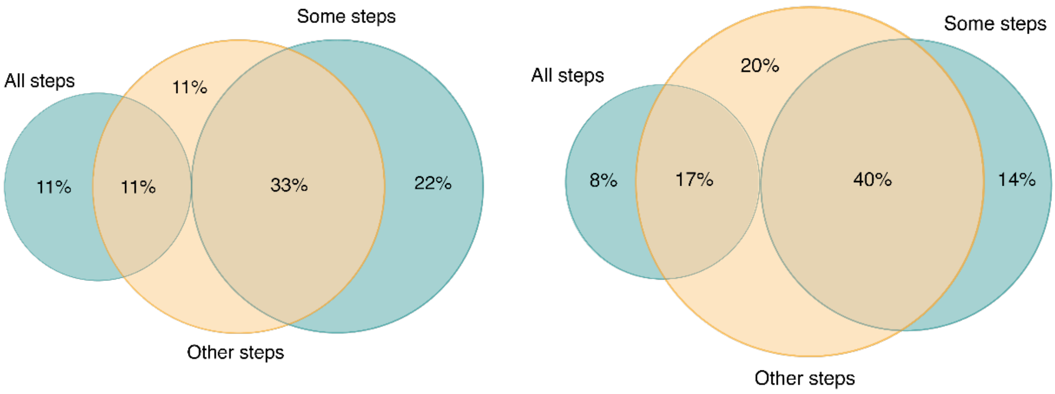


Figure 23 Overview performed actions in e-mail (left) and walled garden (right) group

7.6 Sub-conclusions on cleanup efforts

'What actions do Mirai-infected consumers perform?' is the central question in this chapter. We conclude that the majority of consumers do not follow the recommendations in the notification. When looking purely to compliance with all remediation steps as recommended in the notifications, we can conclude that consumers who actively perform actions score badly: respectively 22% and 25% of consumers with an e-mail and in walled garden performed all steps. Striking in these findings is the high percentage of active and complying consumers in the e-mail group. Since the incentive is less than for consumers in the walled garden, one could have expected lower rates. The disconnection from the Internet is thus not the only motivation to comply with a notification.

Another peculiarity is the number of actions that are performed that were not mentioned in the notifications. Consumers in the walled garden particularly take more drastic measures such as disconnecting the device from the Internet (23%), or discarding the infected device completely (20%). Despite the non-compliance among consumers who received a notification, some of them have performed actions that remediate Mirai. For example, the disconnection of devices and update of Domoticz software lead to successful remediation. This means that we cannot purely determine remediation based on compliance, but rather on the specific actions a consumer took.

In the control group, we find no clear explanation for remediation. Ten of the 45 interviewed consumers cleaned up Mirai unintentionally by updating outdated software; one consumer by the disconnection of devices. Although some of these consumers experienced troubles with their device, none of them was aware of the fact that his/her device was infected with Mirai. One must note that the consumers in the control group who unintentionally cleaned up their device mainly exist of consumers with a device running on outdated Domoticz software. Only two consumers cleaned up different device types. Without the outbreak of the Domoticz-variant, the share of consumers who cleaned up their device would have been lower. Despite these cleanup efforts, there is still a difference of 40% difference in remediation after two weeks which we cannot explain. Although the interviews may have stimulated remediation, we don't detect such influence when comparing the interviewed control group with the non-interviewed control group.

Remarkably, almost all consumers – including those in the control group – were able to identify IoT devices in their home. The ease in identifying the infected device varied among consumers. Many consumers who received a notification needed additional help to find out which of the devices would be infected. Their requests differed from specific inquiries ('which Mac address') to general questions ('how do I find the device if I cannot install a virus scan on it?'). However, some consumers already had a gut feeling of which device was infected ('that Chinese cheap camera', 'that DVR that was already malfunctioning').

In conclusion, the appeal to consumers to remediate Mirai seems effective when looking at the numbers of performed actions: 98 % of the walled garden performed at least one action, versus 92 % in the e-mail group. In contrast, 88% of the consumers among the control group didn't perform any cleanup actions. The 22% who did perform actions are mainly consumers who own a device running on Domoticz who updated their software.

These sub-conclusions are illustrated in figure 24. The described cleanup efforts are part of the 'behavior' phase of the theoretical framework. We can now further specify this phase by distinguishing intentional and intentional cleanup efforts. One out of five consumers in the control group unintentionally cleaned up Mirai. Notified consumers intentionally performed cleanup actions but most of these didn't comply fully to the recommended actions. We cannot point out the exact reason for this. One explanation is the lack of good comprehension. Since many consumers asked for additional help, we conclude that consumers appear don't have a full understanding of how to tackle the problem. This may obstruct compliance. Another explanation is the lack of motivation to comply with all recommendations. We observed no pronounced cases in which consumers didn't intend to comply due to a lack of faith in their capabilities ('self-efficacy') or the effectiveness of the recommended actions ('response-efficacy'). However, the high rate of full compliance indicates that either consumers may not fully rely on the advice and prefer to solve the problem in their own way, or consumers believe they are not capable of performing the actions and thus take rigorous actions.

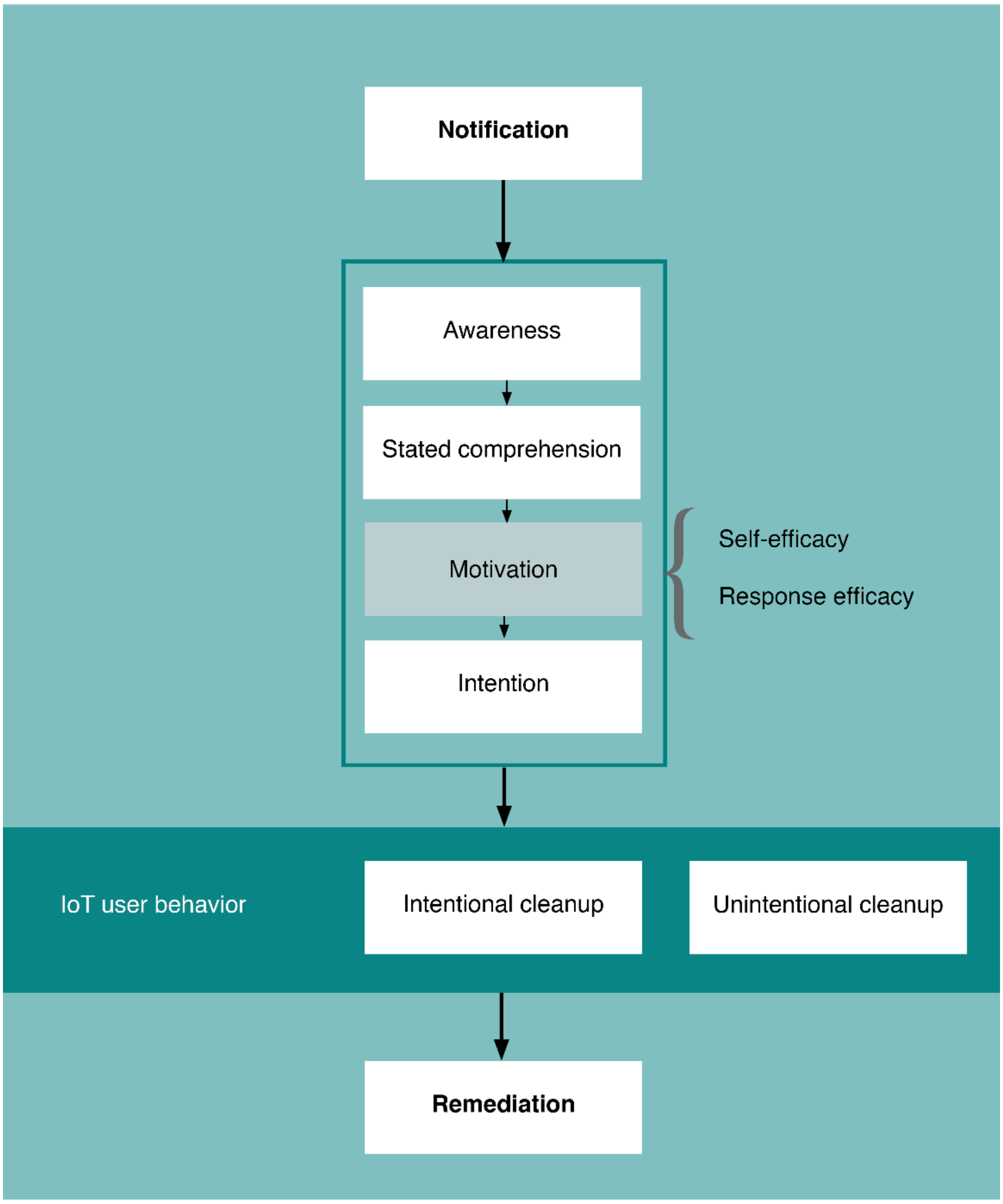


Figure 24 Theoretical framework adjusted to sub-conclusions on cleanup efforts

8 Compliance and remediation

8.1 Introduction

The first goal of this chapter is to understand what obstructs compliance with the recommended cleanup actions among notified consumers. This is formulated in the research question 'What are the reasons for non-compliance with Mirai notifications?'. In the previous chapter, we concluded that the majority of consumers does not comply with the recommended cleanup actions, but that still many consumers took other measures that are effective for Mirai cleanup. Therefore, we not exclusively look to 'strict' compliance, but rather to performance of right cleanup measures (a looser notion of compliance). In section 8.2, we analyze the reasons for compliance and non-compliance (both in loose sense), using the theoretical framework.

Secondly, we use the information we have about consumers' behavior to make the first step in exploring its effect on remediation. This helps in partly answering the research question 'How can remediation of Mirai-like bots be explained?'. Section 8.3 presents an exploration of the effect of performing right cleanup actions on survival behavior. Section 8.4 presents the drawn conclusions.

8.2 Dissection of reasons for (non-)compliance

The theoretical framework, as presented in section 2.4.5, describes five phases that lie between notification and desired behavior. Each stage is a potential obstruction in the way of achieving remediation. For both e-mail and walled garden notifications, we mapped the paths between these phases. Each node presents one stage of the theoretical framework. Since we are interested in the loose notion of compliance (whether a consumer has cleaned up Mirai), we exclude compliance in a strict sense. This results in the following four mapped stages:

- Awareness: whether a consumer has received a notification. This is not only technical delivery (correct e-mail address) but also includes awareness of the consumer about the notification (e.g. not regarded as a spam e-mail);
- Comprehension: whether the content of the notification was clear to the recipient;
- Intention: whether a consumer had the intention to comply with the notification. In other words: was s/he motivated?;
- Correct measures: based on the findings in the previous chapter, we choose here to distinguish compliance from right behavior. Whether a consumer has performed right cleanup measures is determined by the following rules:
    - The consumer has restarted the infected device and changed the password. Or:

- The consumer has disconnected the device from the Internet. Or:
- The consumer has discarded the device. Or:
- The consumer has updated/reinstalled Domoticz software.

The tree with reasons of (non-) compliance is visualized in figure 25 for the consumers who are placed in the walled garden. These consumers in a walled garden were almost all aware of the notification. For 68% of these consumers, the content was clear and intention to comply was present. Despite the unclarity for the 32% other consumers, the majority (92%) still intended to try to comply with the remediation steps. 31 of the 39 (79%) consumers in walled garden succeeded in performing correct cleanup actions.

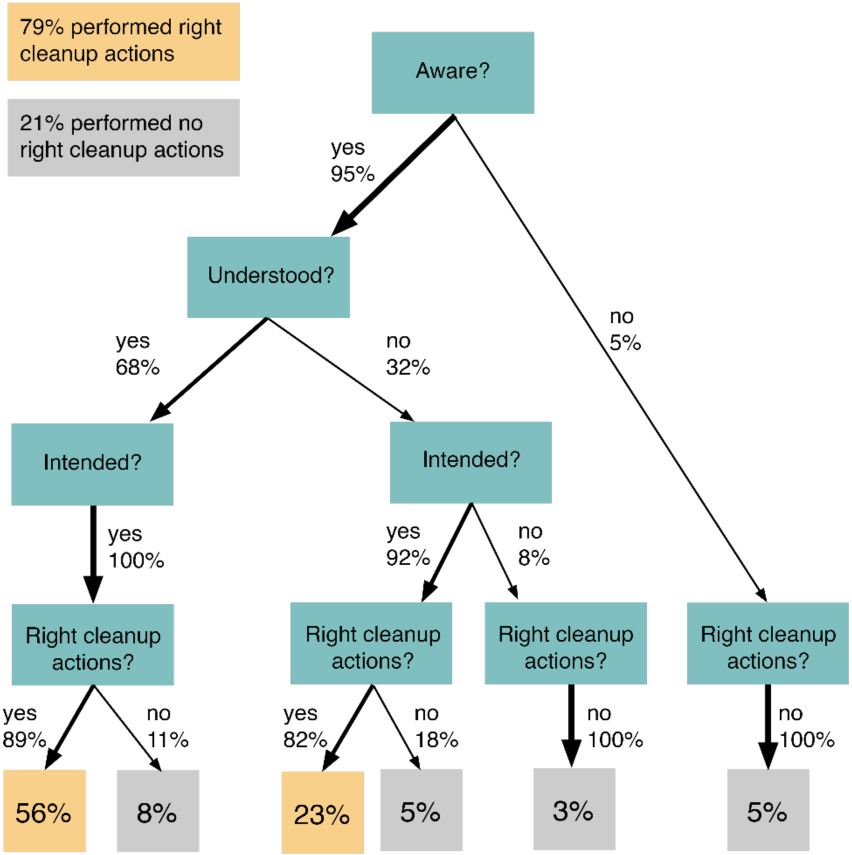


Figure 25 Reasons for (non-)compliance with walled garden notification

Figure 26 shows the tree for consumers who only received an e-mail notification. In comparison, these consumers were less aware of the notification: only 75% of the consumers read the notifications. Not all consumers who understood the message were motivated to comply. Of all people who did understand the content and intended to comply, seven succeeded in performing correct actions (58% of total). Of all consumers who didn't read the notification, still one of the three remediated Mirai.

Both figures 25 and 26 show that the majority of notified consumers stated the notification was clear to them. However, the findings of chapter 7 contradict this: many consumers needed additional help.

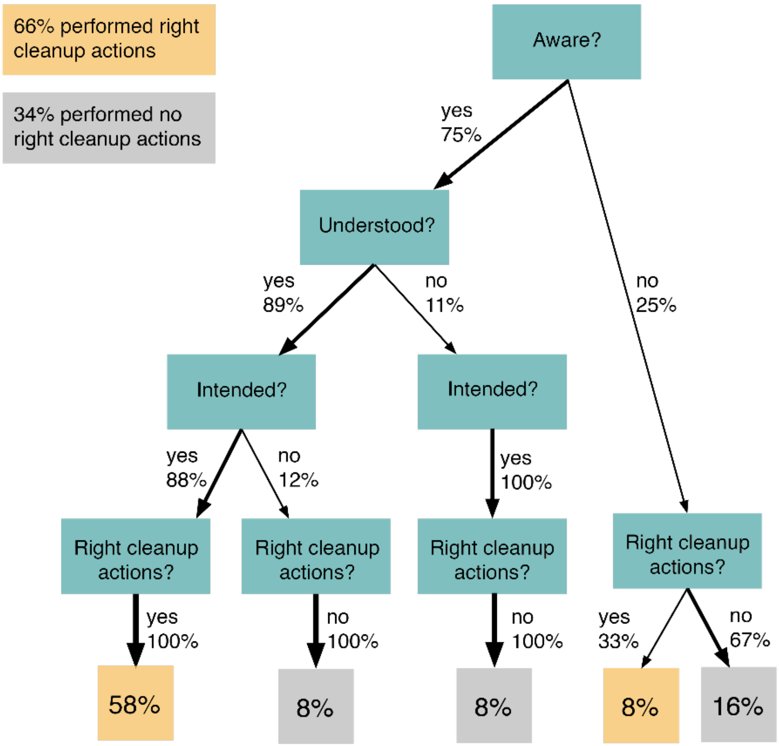


Figure 26 Reasons of (non-)compliance with e-mail notification

Table 7 explores the intention-behavior gap of notified consumers. Behavior is defined in two ways: strict compliance with the notification (performing all recommended steps), and loose compliance (performing correct cleanup actions, can also be others than mentioned in the notification). Concerning strict compliance, only 25% of the consumers who intended to comply did actually do so. When looking purely at performing correct cleanup actions, the gap is smaller: only 14% of the motivated consumers did not succeed in taking the right cleanup measures.

Number of consumers who were aware of the notification	Of who intended to comply:	Of who complied to all actions in the notification:	Of who performed right cleanup actions:
51	44 (86%)	11 (21%)	38 (75%)

Table 7 Exploration intention-behavior gap

Although the majority of consumers who are aware of the notification have the intention to comply, the underlying motivation differs per notification mechanism. The bar plot in figure 27 shows that the disconnect from the Internet in the walled garden is the primary reason to comply, while intention in the e-mail group is mainly driven by the wish for a secure network and Internet. Among the consumers driven by security concerns, the motivations have nuance differences. Some consumers wanted to comply because of concern for the security of their own network and privacy (‘I’m afraid for theft of my personal files on my computer’), other consumers expressed concern for the threat to society in general (‘I don’t want to spread all kinds of viruses to others’, ‘I don’t want to contribute to a DDoS attack’). Other reasons to comply to the recommendations vary from encountered issues (‘my devices were already malfunctioning and my internet was getting slowly’), to one consumer who wanted to avoid a potential walled garden placement KPN. Three consumers express disgust towards the Mirai-infected device ('what a source of misery!', 'I already doubted the device, I brought it straight to the recycling dump','I want to lose that thing, have it off my network!').

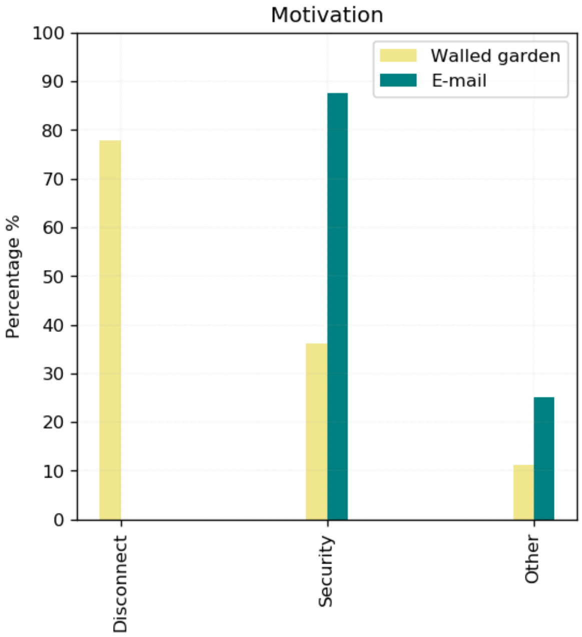


Figure 27 Motivation to comply with notification per notification mechanism

8.3 Exploring the effect of behavior

Figure 28 shows - per treatment group - the survival curves of consumers who performed the right cleanup actions and of those who didn't. The control group has the lowest remediation rate (78%) among consumers still infected after two weeks while having performed the right actions. As described in chapter 7, these actions mainly encompassed the update of outdated Domoticz software. It is remarkable that still two of the nine consumers appear on our radar after the experiment period. When looking at the consumers who didn't clean up their device, 58% is not seen again after the experiment period. Although this rate is 20% lower than the other group, the difference between the curves is non-significant (Log-rank test estimates a p-value of 0,31)[5].

The curves within the e-mail group differ most of all treatment groups. The remediation rate among consumers who didn't perform right clean up actions is the lowest in this group with a rate of 50%. One of the eight consumers who remediated is still infected after two weeks. Although the remediation rate of the two groups differs with 37% after two weeks, the Log-rank test estimates no significant difference (p=0,18)[5].

The walled garden group contains the lowest remediation rate for both the consumers who performed the right actions and the consumers who did not. Six of the eight consumers who did not perform any right cleanup action are not detected after three days. Only 10% of the consumers who did perform the right actions are still infected after the experiment period. Similar to the other treatment groups, there is no estimated difference between the two survival curves (p=0,27)[5].

[5] As explained in section 4.5.1, due to omitted-variable bias we may not estimate a significant difference between survival curves while the influence of behavior variable could be significant.

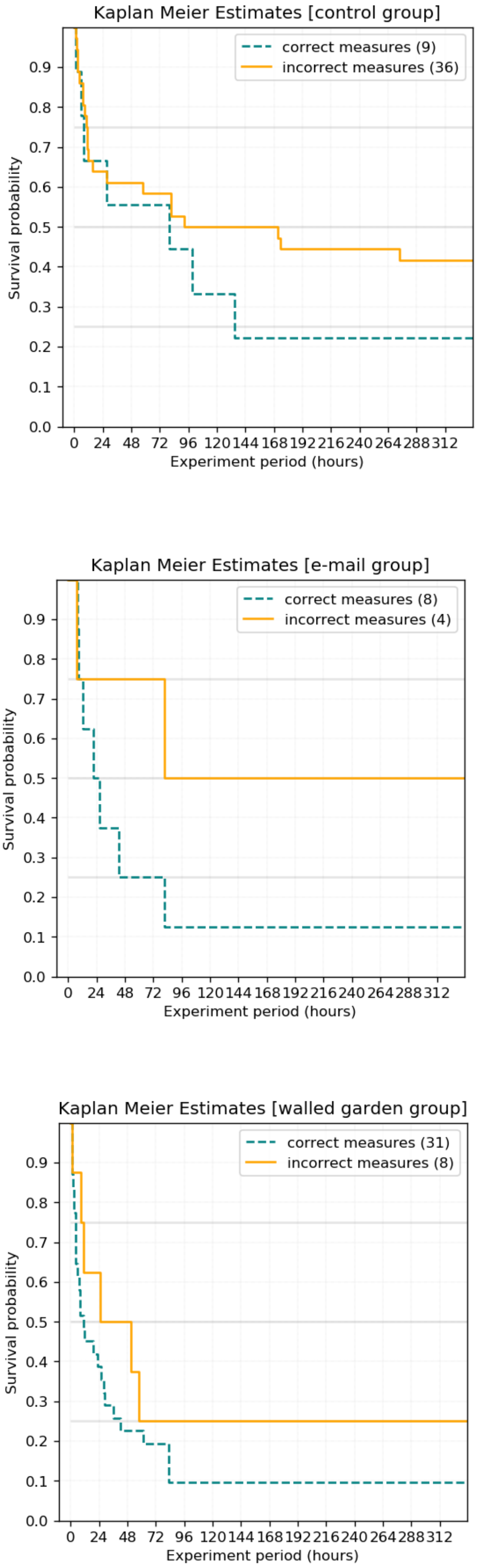


Figure 28 Survival curves - the influence of behavior per treatment group

8.4 Sub-conclusions on compliance and remediation

The first research question focused on in this chapter is ‘What are the reasons for non-compliance with IoT abuse notifications?’. Both e-mail and walled garden notifications are effective in reaching the consumer, informing them and encouraging them to take action. Most consumers who are placed in a walled garden have as a primary incentive to get back Internet access while people who received an e-mail are motivated by the severity of the threat. Note that the e-mail notifications were only sent to Telfort consumers and that the findings on the effect of e-mail only apply to this market. When reviewing these findings in light of the theoretical framework, the identified motivations are covered by two motivation components of the theoretical framework. The disconnect from Internet access can be considered a reversed ‘response cost’: instead of the costs to clean up a device, it is the costs of not doing so. The other common motivation - the whish for a secure network – can be traced back to the motivation component ‘perceived severity’. People are motivated because they believe a Mirai infection is a severe problem for themselves and/or society.

Although notifications are effective in reaching, informing and activating consumers, we identify a large intention-behavior gap. Only 25% of the consumers who state to be motivated to comply, succeed in doing so completely. On the other hand, looking at compliance in loose sense (taking effective measures), the intention-behavior gap is smaller: 14% of the consumers did not manage to clean up their infected device. In addition to the intention-behavior gap, we observed a discrepancy between the stated comprehension and observed comprehension. Although the majority of consumers stated that they understood the content of the notification completely, many consumers were not able to clean up their infected device without additional help.

The second goal of this chapter was to make a start with answering the research questions ‘How can remediation of Mirai-like bots be explained?’. Despite our attempt to explain natural remediation by looking at user behavior, we are (still) not able to do so. We observe a substantial remediation rate among consumers who didn’t clean up their device. The unexplained remediation is highest among consumers in the walled garden group: 75% of the consumer who didn’t clean up their device is observed as remediated during the experiment. The lowest remediation rate is among the e-mail group: half of the Mirai infections are observed as remediated. Since a Mirai infection cannot disappear without reason, we must look for other explanations for the observed remediation. The first explanation is that we have not observed all behavior. We provide four possible scenarios:

- Consumers forgot what clean up actions they performed or forgot to mention them during the interviews;
- Consumers clean up their device unintentionally;
- The device is cleaned up by someone else in the household without the consumer’s knowledge;
- The consumer cleaned up the device after the interview: between the end of the experiment period and the end of the two-week additional observation.

The second explanation concerns the area we focus on. This study studies remediation from the user perspective. Since the unexplained natural remediation rates are so high, one can argue that unobserved behavior is not able to explain that complete gap. That would imply that more than half of all consumers cleaned up their device while stating otherwise during the interviews. Since we believe this is unlikely, we must look for answers on the attacker side. We provide the following two unexplored explanations:

- Other malware takes over Mirai-infected devices. This study focuses on Mirai and only included abuse feeds on this malware. Therefore, a device may be compromised by other malware without our knowledge. That would explain why an IP address doesn't appear on our radar and the bot is wrongly considered to be cleaned up.
- The majority of Mirai infections are detected when a bot is in a scanning phase. Conventional Mirai bots are scanning unless they get commands from the botnet herder. With the increasing number of Mirai-variants, bots may have evolved scanning behavior. They might for example only scan when commanded to, or have built-in behavior that determines that a bot only scans in the first hours of its life.

The findings of this chapter are illustrated in figure 29. Although we didn't estimate a significant difference between the survival curves of consumers who said to have cleaned up their device and those who didn't, we cannot conclude yet that the observed behavior has no influence on the estimated infection time. This has to do with the omitted-variable bias as explained in section 4.5.1 and 6.5. Especially now we know that we didn't observe all behavior of consumers and potential influences from the attacker side, the univariate Kaplan-Meier survival cannot provide us with a definitive conclusion on the effect of stated behavior. The models in chapter 10 include more variables and thus provide us with more reliable estimates.

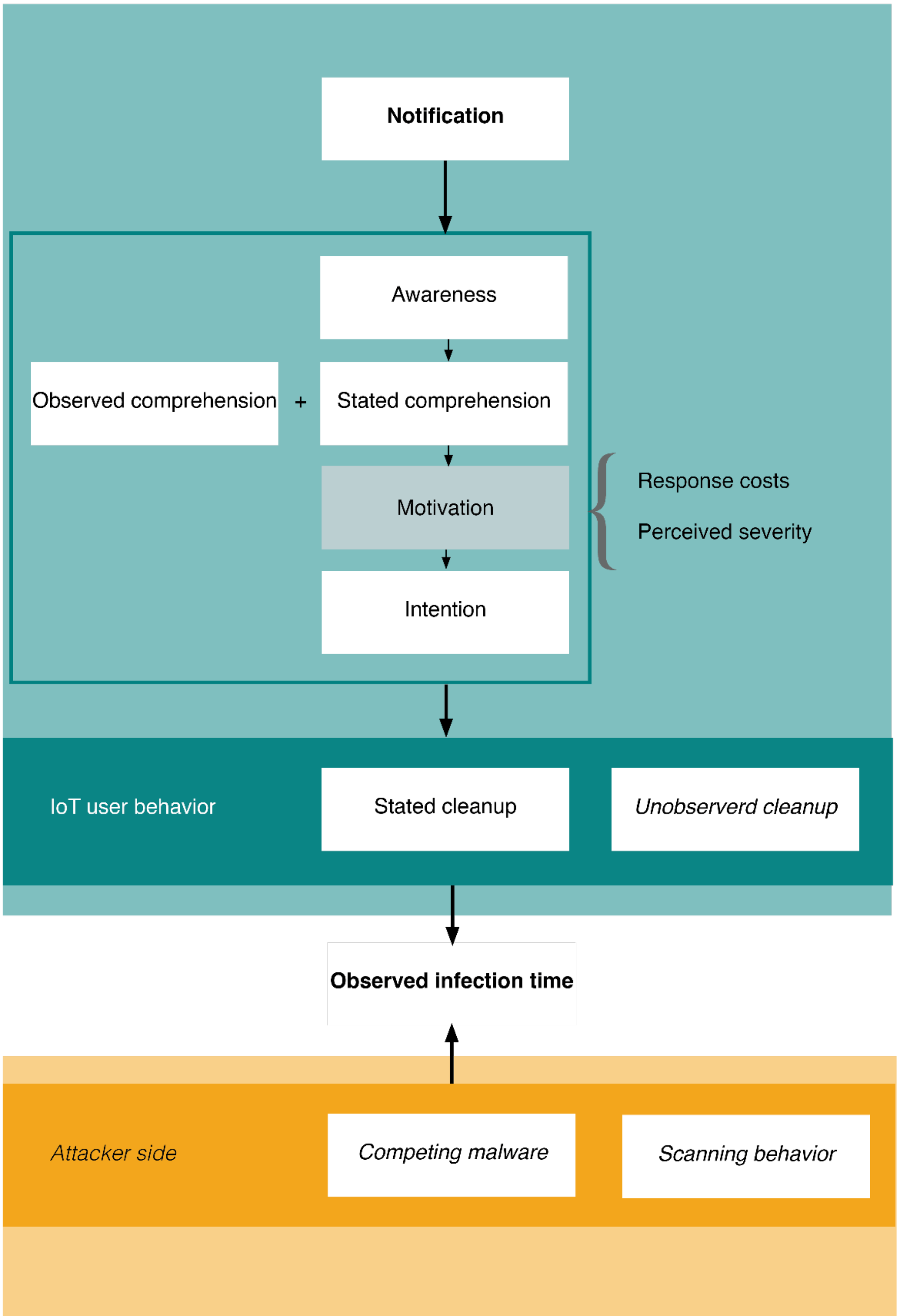


Figure 29 Theoretical framework adjusted to sub-conclusions on compliance and remediation

9 Customer experience

9.1 Introduction

The previous two chapters focused on consumers' behavior after receiving a notification, what obstructed them in complying and what motivated them to clean up their Mirai-infected device. This chapter covers customer experience with as central research question: 'How do consumers experience Mirai notifications?'. Section 9.2 presents the result on customer satisfaction. The obtained suggestions are described in section 9.3. Section 9.4 presents the conclusions.

9.2 Customer satisfaction

All interviewed consumers who received notification were asked about their experience. Figure 30 displays two pie charts of consumer satisfaction of consumers who were placed in a walled garden (left) versus consumers who received an e-mail (right). 61% of the interviewed consumers in the walled garden group were satisfied versus 100 % in the e-mail group. Almost a quarter of the consumers placed in a walled garden were dissatisfied. These results are contradictory to Altena's (2018) study in which 8% of all notified consumers expresses satisfaction. The only explanation we can find is the difference in the interview protocol. In Altena's (2018) study, consumers are not asked specifically about their opinion, only about their suggestions ('How could the communication to customers be improved when KPN sees problems like this?'). In our study, we asked specifically about a consumer's opinion ('What do you think of KPN's service to reach out to infected customers?'). Apparently, consumers only express their satisfaction when asked about. This is also reflected in the communication with the Abuse Desks: the contact forms and e-mails are often framed negatively ('I want my Internet back soon!') while most of these consumers are actually grateful for KPN's service.

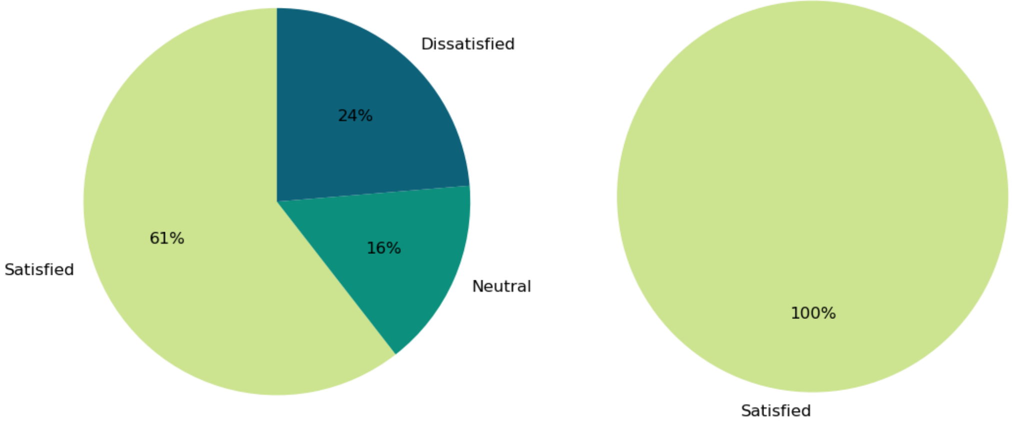


Figure 30 Consumer satisfaction for walled garden (left) and e-mail (right) notifications

Table 8 contains measures of central tendency and figure 31 present the survival curves of consumers with a different experience. Satisfied consumers in our experiment have a 'typical' infection time of 12 hours, which is half the infection time of dissatisfied consumers and 30 hours less than consumers without a clear opinion about the notification service. In contrast to these differences, we observe no notable differences in survival behavior in figure 31 (log-rank test estimates p-value higher than 0,31 for comparison of all curves).

	mean	std	median
Satisfied	50,53	95,46	11,80
Neutral	88,52	124,80	42,04
Dissatisfied	61,33	107,06	24,92

Table 8 Descriptives on infection time (in hours) based on consumer satisfaction

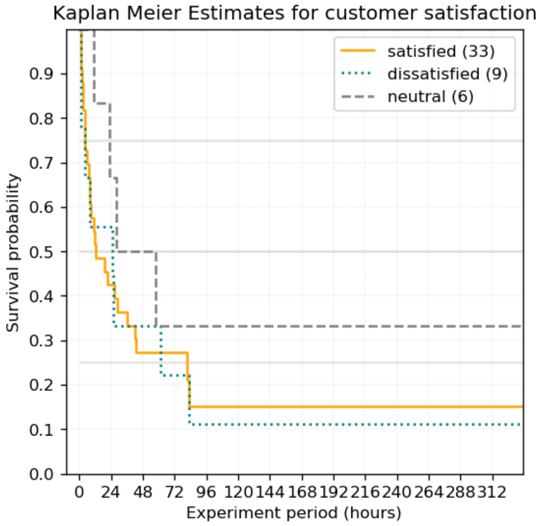


Figure 31 Survival curves - customer satisfaction

9.3 Customer suggestions

Consumers who have fallen victim to a Mirai infection may possess valuable information to improve notification. They may point us to blind spots and may have creative ideas for improvement. Of all notified consumers, 27 (69%) of the walled warden group and six (54%) of the e-mail group had suggestions on how to improve notifications concerning Mirai. These suggestions can be divided into two categories: notification content, and notification procedure.

9.3.1 Notification content

Eight consumers declare that they questioned the authenticity of the notification and suggested this may be improved. Most of these people suggest a more personalized content. One person specifically says it would be helpful if he could have verified the authenticity of the message. Two customers wonder why they have never heard of the Abuse Desk before and suggest better publicity of this service.

Five consumers believe that the measures are not easy to follow for technical lay(wo)men. For example, one consumer had difficulties in regaining Internet again after resetting his modem.

It took him long to realize the default password is written underneath the modem and he suggested to include such little tips. Eight people specifically suggested more help or advice on how to detect the infected machine (‘isn’t there any tool available?’). The people asking for this differed from technical experts (who asked for a MAC address) to people who did not manage to identify one single Internet-connected device and wanted a list of device types to understand what kind of device he was looking for.

9.3.2 Notification procedure

The suggestions about the notification procedure only come from customers who were placed in a walled garden. All consumers who were only e-mailed were satisfied with how they were approached.

Twelve of the consumers in the walled garden group would have liked to receive a notification before being placed in quarantine. Many of those consumers encountered the problem that they were not aware of the quarantine until very late (because they did not receive a landing page in their browser or because they did not go to their inbox because they believed their Internet was down). Five consumers suggested an additional call for notification purposes.

Quite many people did not manage to release themselves (44 consumers in total) due to the absence of a landing page. Six consumers suggest clear information on how to get released after taking remediation actions. Five people express their concern about the disconnect from the Internet. Two of those were sons (both adults) of consumers and were worried about the Internet disconnect of their parents. One of them mentioned that the emergency button would not function. The other explained that his parents were very worried because the surveillance camera did not function anymore. One consumer has been robbed during the disconnect from the Internet since he had no functioning surveillance camera.

Seven consumers wanted to have been able to call the Abuse Desk for additional help. Three consumers mention that the Help Desk has not been able to help them. Two consumers expressed their dissatisfaction with the limited availability of the Abuse Desk. Both men could not release themselves and had to wait until the following day before they were manually released by an Abuse Desk employee.

9.4 Sub-conclusions on customer experience

The central research question of this chapter was ‘How do consumers experience Mirai notifications?’. In general, notified consumers are satisfied with KPN’s effort to notify them about Mirai while in most cases this was not reflected in the communication with the Abuse Desk. A quarter of consumers placed in a walled garden are dissatisfied which can be explained by the disruptive nature of the measure. All consumers who received an e-mail are satisfied with this service. Customer satisfaction does not seem to have a significant effect on the estimated infection time.

The suggestions of consumers vary greatly but three things stand out:

- People wish to be better informed. This is the case for both consumers who have difficulty with following the steps, as for more tech-savvy consumers who wish to have more details on the information that KPN possesses about the abuse incident. This finding is in line with the findings of chapter 7 in which we observed that many consumers requested additional help from the Help Desk and Abuse Desk.
- In addition to the first point, consumers would like to have additional help from KPN employees. The Abuse Desk cannot be called and employees of the Help Desk often know

little about the abuse incident or are not able to help a consumer. People wish to be able to call someone for help on the problem or for a release from the walled garden. In addition, consumers wish to ask for additional help outside office hours so they do not have to wait a night or weekend without access to the Internet.

- On top of the disruptive nature of a quarantine environment, three factors aggravate the disconnection: A) consumers are not aware of the walled garden up until they come home in the evening when the Abuse Desk is already closed and not able to help. B) The majority of consumers is not able to release themselves because they have not received a landing page, accidentally clicked it away or are not aware of the existence of this page. C) Consumers think that because their Internet is down, they cannot enter their e-mail inbox. This prevents them from retrieving the information that tells them they are put in a walled garden and explains them how to self-release. Therefore, it takes a long time for some consumers to find out they have to take action. A warning prior to an Internet disconnection is, therefore, a frequently mentioned suggestion.

10 Remediation drivers

10.1 Introduction

Previous chapters have reported the results and conclusions on user characteristics (chapter 5), Mirai remediation among different populations (chapter 6), cleanup efforts (chapter 7), reasons of non-compliance (chapter 8) and customer experience (chapter 9). This chapter combines these aspects to explore what factors influence remediation. The results give an answer to the research question: 'How can remediation of Mirai-like bots be explained?'. The data is analyzed using the Cox and AFT modeling techniques which are both described in chapter 4. The modeling is done in three steps as illustrated in figure 32 - from generic to specific. Each step is reported in a separate section:

- Step 1 (section 10.2): includes the observations of all consumers, using only information from subscription accounts (e.g., gender) and variables as set by the experiment (e.g., treatment);
- Step 2 (section 10.3): includes observations of only interviewed consumers, using additional variables obtained from interviews (e.g., device type and cleanup efforts)
- Step 3 (section 10.4): includes observations of only interviewed consumers who have received a notification, using additional variables obtained from the interviews (e.g., comprehension of the content and intention to comply).

Section 7.5 draws conclusion using the findings from the models. Appendices J, K and L contain the detailed modeling process of respectively section 7.2, 7.3 and 7.4.

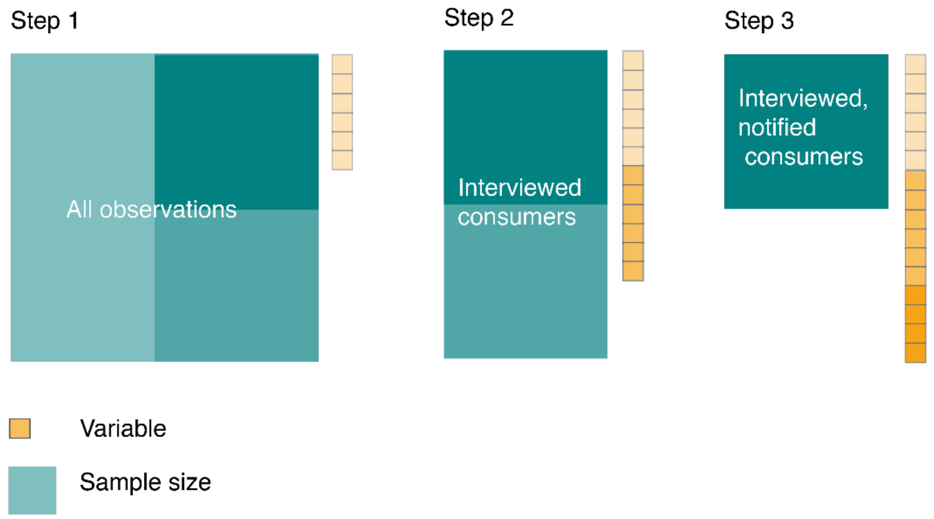


Figure 32 Three modeling steps

10.2 General exploration remediation drivers

10.2.1 Variables of interest

To be able to include all observations in the modeling process, we can only use variables which are known for all observations. In this first modeling step, all the variables that are addressed in chapter 5 on study population are included, in addition to the variable 'time-splits' which refers to the division of observations before and after the Domoticz-variant outbreak. Table 9 provides an overview of these variables, including the dependent variables. Appendix J provides more details on how the dummy variables are coded. The dataset contains 177 observations, of which 28 censored (15%). We test whether the independent variables are covariates for infection time. In other words: we explore what variables have a significant influence on remediation and what these relations look like.

	Variable	Explanation	Coded
Dependent variables	Infection time	Infection time of a Mirai bot	Between 0 and 336 hours
	Death	Censored observations: all the bots that were still infected after two weeks.	Censored = 0 (still infected) Not censored = 1 (last detection is within two weeks after notification)
Independent variables	Sex	Male	Two dummy variables
		Female	
		Business	
	Age	2019 minus birth year	Continuous variable
	Market	Telfort or KPN consumer	Dummy variable 0 = KPN 1 = Telfort
	Treatment	E-mail notification	Two dummy variables
		Walled garden	
		No notification (control group)	
	Time splits	Whether the infection took place before or after June 9th	Dummy variable 0 = before 1 = after

Table 9 Variables included in the first modeling step

10.2.2 Modeling 177 observations and 5 variables

Modeling a Cox hazard model (see appendix J.3) based on the data and variables as described in the previous section leads to a bivariate model wherein the variables 'female' and 'walled garden' have a significant effect on the infection time. The dummy variable 'female' refers to the distinction between female consumers (coded as 1) and the rest of the consumers (male + unknown gender, coded as 0). The coefficient of this variable in the partial hazard is 0,52[6]

[6] Exp(0,52)=1,68

(CI:0,03-1,00, p=0,04), which indicates a factor 68% increase of baseline hazard. In other words: female consumers have about 70% more chance on remediation compared to consumers who are not known to be female. The 'walled garden' variable distinguishes consumers who are placed in a walled garden versus the other consumers (control and e-mail group). The coefficient of this variable in the partial hazard is 0,57[7] (CI:0,19-0,95, p<0,005), which implies a 77% increase of baseline hazard. Combined, the estimated regression equation is:

$$\log\left(\frac{group\ hazard}{baseline\ hazard}\right) = 0{,}52x_{female} + 0{,}57x_{walled\ garden}$$

Wherein $x_{female}$is an indicator for gender (1=female, 0=rest) and $x_{walled\ garden}$ is an indicator for notification mechanism (1=walled garden notification, 0=no notification or e-mail). Female consumers placed in the walled garden have the highest relative increase in hazard rate. This group has three times[8] higher remediation rate compared to the baseline hazard.

When modeling an AFT model (see appendix J.4) with the same variables and observation, the LogNormal distribution has the best fit. The AFT LogNormal model that is estimated also results in a bivariate model with 'female' and 'walled garden' as significant covariates. The coefficient of 'female' is -1,03 (CI:-2,07- -0,01, p=0,052) and the coefficient of 'walled garden' is -1,13 (CI:-1,92- -0,34, p=0,005). This means that female consumers have a 68%[9] decrease in mean and median Mirai-infection time and consumers placed in a walled garden a 64%[10] decrease. The influence of these two covariates combined can be described by the accelerated failure rate λ:

$$\lambda(x) = \exp\left(-1{,}03x_{female} - 1{,}13x_{walled\ garden}\right)$$

Female consumers placed in a walled garden have an acceleration rate of 0,12[11]: their mean infection time is 88% shorter than consumers not placed in a walled garden and who are not known to be female.

The influence on remediation of the two estimated covariates are shown in appendices J.3 (Cox model) and J.4 (AFT model). Figure 33 visualizes the *observed* - not modeled - survival curves based on these two variables to gain a better picture of the data from which the models are derived. Since we have not controlled for demographics, consumers of each gender are not equally distributed over the treatment groups. Only four female consumers were placed in a walled garden. These four women have remediated Mirai within three days and stand out compared to the other survival curves. Consumers who are not known to be female and are not in walled harden have te lowest remediation rate.

[7] Exp(0,57)=1,77

[8] Exp(0,52+0,57) = 2,97

[9] Exp(-1,13) = 0,32 // 1-0,32 = 0,68

[10] Exp(-1,03) = 0,36 // 1-0,36 = 0,64

[11] Exp(-2,16) = 0,12

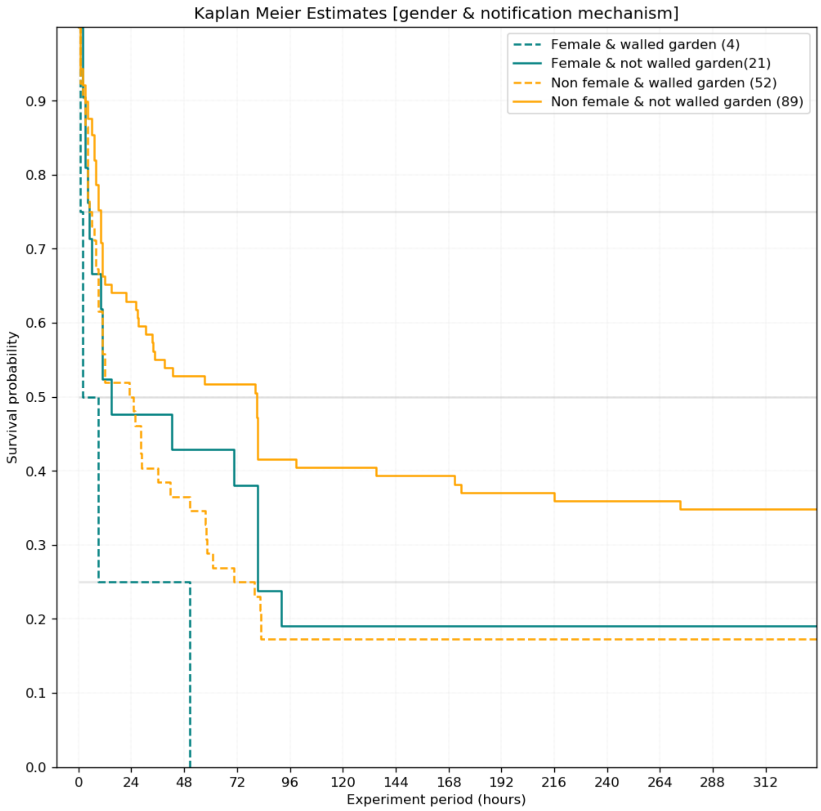


Figure 33 Survival curves - covariates gender and notification mechanism

10.3 Exploration interviewed consumers

10.3.1 Variables of interest

We can use information obtained from the interviews to further specify the drivers of remediation. We include two qualitative variables concerning the stages of the theoretical framework: awareness and behavior. 'Aware of notification' is a variable that distinguishes whether the consumer received a notification *and* is aware of its content. 'Right measures' is a variable that refers to whether consumers have performed effective remediation measures. 'Compliance' and 'Intention' are not included since this only concerns notified consumers. These stages will be explored in section 10.4.

In addition to awareness and behavior, device types are also included in the modeling sequence. The pie chart in figure 15 (chapter 5) visualizes the ratios of device types. However, these pie charts do not take into account that some consumers have identified several devices as possible infected. Since we don't know the precise device, we collect these devices under the variable 'multiple'. The NAS and Rasberry Pi devices are collected under the variable 'home automation' (modeled as 'home').

In total, the dataset of interviewed consumers contains 39 observations of home automation devices, 23 cameras, 11 instances of multiple identified devices, 11 unknown devices, 3 printers, and 2 routers. Due to their low occurrence, printers and routers are not modeled as separate variables but as a residual category. The five device categories are modeled using four dummy variables (see appendix K.1 for the coding schemes).

The six discussed variables are summarized in table 10 and are added to the five variables that are described in section 10.2. From the observations, we exclude consumers who are reinfected and who are not interviewed. The remaining dataset contains 89 observations, of which 22 are censored (25%).

Variable	Explanation	Coded
Awareness	Whether the consumer received and read a notification	Dummy variable 0 = not aware (incl. control group) 1 = aware
Behavior	Whether the consumer performed right cleanup actions (can be other actions than recommended)	Dummy variable 0 = incorrect actions (incl. inaction) 1 = correct actions
Device type	'home': Consumers who have identified a NAS or Rasberry Pi as the infected device	Four dummy variables
	'camera': Consumers who have identified an IP camera as the infected device	
	'multiple': Consumers who have identified multiple IoT devices as possibly infected	
	'unknown': Consumers who were not able to identify an IoT device in their network	
	Consumers who have identified a printer or router as the infected device (residual category)	

Table 10 Additional variables included in the second modeling step

When plotting the observations against infection time (see appendix K.2), we detect that the majority of censored observations are among consumers who have not performed effective measures and among consumers who are not aware of the notification. Appendix K.2 presents a correlation matrix of all variables, which shows that the correlation between these two variables is 0,6. This value indicates that there is a moderate/strong linear relationship between awareness and cleanup actions. Both variables are also correlated with the walled garden variable; awareness (coefficient of 0,75) more than behavior (coefficient of 0,45).

When plotting the observations in the context of device categories, there are no observable trends. The time-splits variable relates to categories of home automation (coefficient of 0,74) and camera (coefficient of -0,6), which is expected. The home automation variable has a moderate positive relationship with behavior (coefficient of 0,45).

10.3.2 Modeling 89 observations and 11 variables

The Cox modeling steps are reported in appendix K.3. This results in a model with three significant covariates: 'female' with a coefficient of 1,35 (CI:0,52-2,17, p<0,005), 'aware fo notification' with a coefficient of 0,64 (CI:0,01-1,27, p=0,05), and 'right measures' with a coefficient of 0,63 (CI:-0,01- 1,28, p=0,05). This leads to the following equation:

$$\log\left(\frac{group\ hazard}{baseline\ hazard}\right) = 1{,}35x_{female} + 0{,}64x_{aware} + 0{,}63x_{right\ measures}$$

Individually, these variables have an increase of the baseline hazard of 280%[12] (female), 90%[13] (aware of notification) and 89%[14] (right measures). Remarkably, awareness of the Mirai infection and right behavior have individually significant explanatory value. Since the 'walled garden' variable showed a moderate positive relationship with both variables, we also estimated a model substituting 'right measures' and 'aware of notification' with 'walled garden'. Although this leads to an accepted model, the model is of less quality (the Akaike information criterion is higher, see appendix K.3).

Fitting an AFT model leads to a bivariate model based on the LogNormal distribution. The modeling steps are reported in appendix K.4. 'Aware of notification' and 'female' are the variables with explanatory value and have a coefficient of respectively -1,78 (CI:-2,78— -0,78, p<0,0005) and -1,96 (CI:-3,565- -0,36, p=0,017). This means that female consumers have a 86%[15] shorter infection time than consumers who are not known to be female. Consumers who have read a notification regarding Mirai have an 83%[16] shorter mean infection time. This variable can be substituted with the variable 'right measures' or 'walled garden' which both lead to an accepted model but with less goodness of fit. The combined acceleration factor can be estimated using the following equation:

$$\lambda(x) = \exp\,(-1{,}96x_{female} - 1{,}78x_{aware})$$

Female consumers who have received a notification and are aware of it have thus the shortest infection time. The mean infection time of this group is 98%[17] shorter than consumers that are not to be known to be female and are not aware of a notification.

In contrast to the Cox model, the LogNormal AFT model does not estimate the 'right measures' variable as a significant covariate (appendix K.4 shows that exclusion of this variable results in a slightly better model). To explore the survival behavior in the experiment, we therefore first look to the combined effect of gender and awareness. The survival curves are illustrated in figure 34. All female consumers within this dataset have remediated their device within four days, while a substantial share of non-female consumers are still infected after two weeks. However, also the non-female consumers who are aware of the notification have either remediated within four days or not at all (the survival curve is horizontal after four days). It is remarkable that the subscribers who have not received a Mirai notification, still show a high rate of remediation: 100% among female consumers, and 58% among non-female consumers. This is in line with the unexplained natural remediation within the control group.

[12] Exp(1,35) = 3,84

[13] Exp(0,64) = 1,90

[14] Exp(0,63) = 1,89

[15] Exp(-19,61) = 0,141 // 1-0,141 = 0,86

[16] Exp(-1,780 = 0,17 // 1-0,17 = 0,83

[17] Exp(-1,96-1,78) = 0,02 // 1-0,02 = 0,98

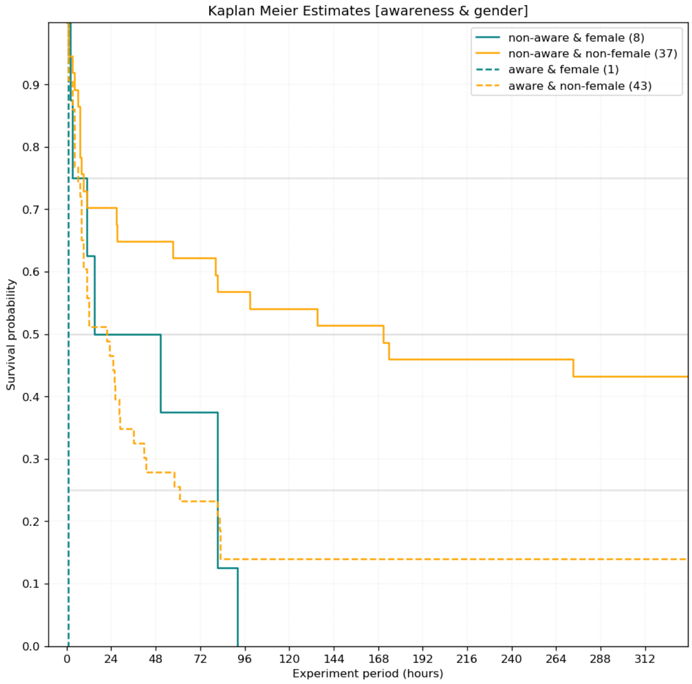


Figure 34 Survival curves - covariates gender and awareness of notification

Of all observations in the experiment, 36% of the female consumers are interviewed (versus 56% of the rest). Since we found in section 10.2 that gender explains remediation, the low amount of interviewed female consumers can obstruct reliable estimates. There are only nine observations of female subscribers, of whom only one is aware of a Mirai notification. Due to the small size of this group, our estimates are not highly reliable.

In figure 35, one can inspect the observations of consumers who performed (in)correct behavior and were (non-)aware of Mirai. The consumers who were either not aware or did not perform the correct behavior are the minority of the observations: both groups cover 20% of all interviewed consumers.

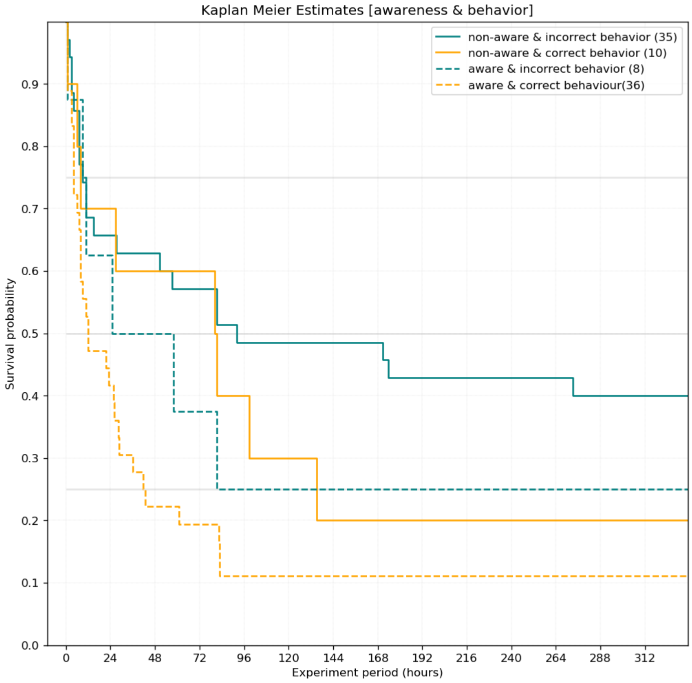


Figure 35 Survival curves - covariates awareness of notification and taking right cleanup measures

None of the device type categories variables are estimated covariates. The survival curves of the four categories are illustrated in figure 36. The behavior of the device categories are quite similar until four days and diverge from there. Infections among consumers with multiple IoT devices identified as possibly infected, have the highest remediation rate. Consumers with an infected IP camera have the lowest remediation rate.

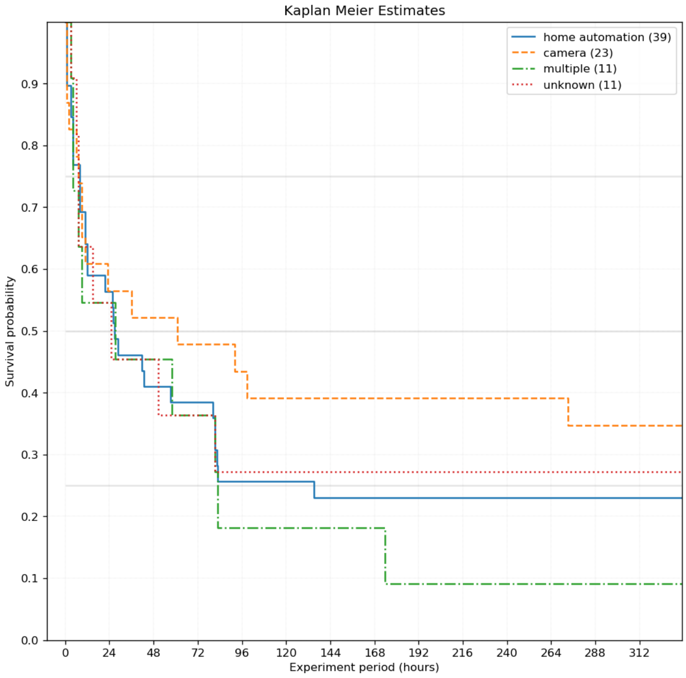


Figure 36 Survival curves – device type categories

10.4 Exploration notified consumers

10.4.1 Variables of interest

In the last step in exploring remediation drivers, we use detailed information that is obtained from the notified consumers who are interviewed. We add four variables to the variables explored before: the two stages of the theoretical framework which were excluded before (comprehension and intention) and two dummy variables regarding customer satisfaction. These variables are explained in table 11. The dummy coding is described in more depth in appendix L.1. These four variables are added to the ten variables that are described in the modeling process of section 7.3 (eleven minus the variable 'e-mail'* since we exclude the control group). In total, the dataset in this modeling sequence contains 49 observations, of which nine censored (18%). Of these 49 interviewed, notified consumers, 37 consumers were placed in a walled garden and twelve received an e-mail only.

Variable	Explanation	Coded
Comprehension	Whether the consumer understood the content of the notification	Dummy variable 0 = not understood 1 = understood
Intention	Whether the consumer intended to comply with the recommended actions	Dummy variable 0 = no intention 1 = intention
Consumer satisfaction	Satisfied with the service	Two dummy variables
	Neutral regarding service	
	Dissatisfied with the service	
Treatment*	Walled garden	Dummy variable 0 = e-mail 1 = walled garden
	E-mail notification	

Table 11 Additional variables included in the third modeling step

When reviewing the correlation matrix (appendix L.3), several correlation coefficients stand out. Firstly, the correlations between all stages of the theoretical framework are in line with our expectations. They all have a positive relationship with each other, with a minimum correlation coefficient of 0,38 (between comprehension and behavior). The intention to comply is strongly (positive) related to awareness (correlation coefficient is 0,83). In a lesser extent, intention also relates to behavior (coefficient is 0,58). Secondly, female subscribers seem to have more trouble with understanding a notification (negative correlation of -0,61) than male subscribes (positive correlation of 0,46). Thirdly, the device type category 'unknown' is negatively related to all stages in the theoretical framework. Lastly, consumers who received an e-mail notification are typically more satisfied with the notification service (coefficient of 0,34), than consumers placed in a walled garden.

10.4.2 Modeling 49 observations and 14 variables

Appendix L.3 reports all modeling steps for the estimated Cox model. The accepted model is a bivariate model with 'male' and 'intended to comply' as significant covariates. 'Male' has a coefficient of -1,22 (CI:-2,21- -0,22, p=0,02) and 'intended to comply' a coefficient of 1,13 (CI:0,07-2,20, p=0,04). Combined the lead to the following equation of the proportional hazard:

$$\log\left(\frac{group\ hazard}{baseline\ hazard}\right) = -1{,}22x_{male} + 1{,}13x_{intention}$$

The male variable has an individual influence of the baseline hazard of -70%[18]. In other words: male consumers have a 70% decrease in the baseline hazard. Note: a decrease in hazard rate implicates a longer infection time. The intention to comply increases the baseline hazard with 211%[19].

[18] Exp(-1,22) = 0,30

[19] Exp(1,13) = 3,11

The variables ‘right measures’ and ‘aware of notification’ can individually substitute the variable ‘intention to comply’. Both substitutes lead to an accepted model, but with decreased goodness of fit and less reliable explanatory value (p=0,06).

When fitting the AFT model, again the LogNormal distribution results in the best fit. All the steps are reported in appendix L.4. The estimated LogNormal AFT model results in a univariate model with ‘intended to comply’ as the only significant covariate with a coefficient of -1,923 (CI:-3,570- -0,276, p=0,22). The failure acceleration factor is:

$$\lambda(x) = \exp\left(-1{,}92 x_{intention}\right)$$

Consumers who have the intention to comply with the recommended steps in the notification thus have an 84 %[20] shorter mean infection time. There is no other variable that can substitute this variable for an accepted model.

The LogNormal AFT model does not include the variable ‘male’. Inclusion of this variable results in an insignificant explanatory value (p=0,055) and a model with a less goodness of fit. The observed survival curves of the combined covariates (intention and gender) are illustrated in figure 37.

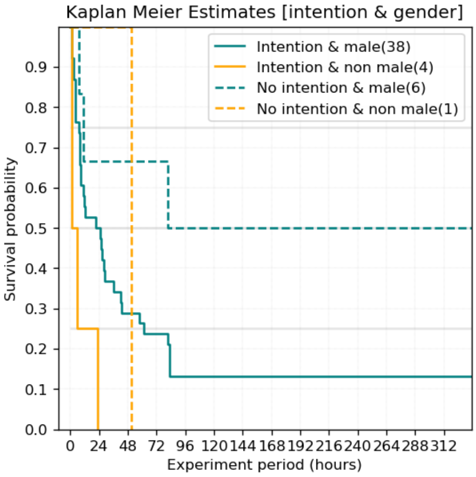


Figure 37 Survival curves – gender & intention

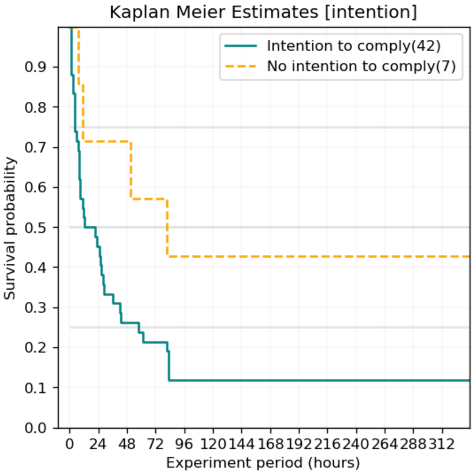


Figure 38 Survival curves - intention

Only five observations in this dataset are of consumers that are not known to be male. Of these five consumers, two are women and three have a shared account. Because of the consequent low reliability of our estimate of the gender variable, we inspect the survival curves with intention as the only variable, which is illustrated in figure 38. The remediation rates after two weeks differ more than 30% between consumers who had the intention to remediate and those who did not.

10.5 Sub-conclusions on remediation drivers

This chapter has provided the results and analysis to answer the research question ‘How can remediation of Mirai-like bots be explained?’. We use three steps of analysis, which enabled us to include all observations in the analysis, as well as all obtained information from the interviews. We used two modeling techniques to analyze the data and thereby identify variables that influence Mirai infection time. Whereas the Cox models are more close to the data (they estimate the relation of a covariate to the observed hazard rates), the AFT models

[20] Exp(-19,23) = 0,146 // 1 – 0,146 = 0,854

provide a more generic view on the influence of a covariate (the failure acceleration factor of a covariate, which implies the relation to the mean infection time).

When including all observations and only the variables known for all observations, we find that gender and walled garden notifications explain remediation. At any given time within the two first weeks of infection, female subscribers have a 68% (CI: 3%-170%) more chance on remediation compared to male subscribers and subscribers of unknown gender, and a 1% to 87% shorter mean infection time. This is a conservative estimate: the other modeling steps estimate a higher influence on remediation but are less reliable due to the small number of female consumers in the datasets.

Consumers placed in a walled garden have 77% (CI:21%-260%) increased chance on remediation compared to consumers in the control and e-mail group, and a 29% to 85% shorter mean infection time. When including the data obtained from interviews, we find that we can further specify the role of walled garden notifications. The Cox model estimates that awareness of a notification (a consumer has received and read the content) and behavior (consumers has performed correct cleanup actions) are of influence. Both variables show a moderate to strong relationship with the walled garden variable. From this relationship, in combination with the finding that awareness and right measures have combined more explanatory value than the variable walled garden alone, we can conclude two things:

- Walled garden notifications are effective because these notification raise awareness and stimulate right cleanup efforts;
- Since awareness of a notification has individual explanatory power - on top of the explanatory power of right behavior - we can conclude we do not observe all cleanup efforts. The reason behind this is the following: the fact that a consumer is aware that his/her IoT device is Mirai-infected, doesn't explain remediation of the device directly. Apparently, the consumer has done something with the device that caused cleanup of the Mirai infection. Since we included behavior in our model and estimated it as a covariate as well, consumers who are aware of the notification thus performed actions which are not included in our data. In other words: there is a discrepancy between *stated* behavior and *actual* behavior.

This finding matches our explanations for natural remediation as described in section 8.4. In the AFT model, exclusion of the behavior variable results in a slightly better model. We can therefore not say that behavior has an effect on the mean infection time (only on baseline hazard).

When modeling the data obtained from notified consumers who are interviewed only, we find that the intention to comply has the best explanatory value. Consumers with the intention to comply with the recommended actions in a notification have 211% increased chance on remediation compared to the baseline hazard, and 84% shorter mean infection time. Since intention has more explanatory power than behavior, the same logic as before applies here: there is unobserved behavior which is influenced by the intention to comply.

In addition to the estimated covariates concerning the stages of the theoretical framework, gender explains remediation as well. The remediation rate among women is higher than among men. Following the logic as applied above, we cannot assume that infected devices are cleaned up just because its user is female. Instead, we conclude that there is unobserved behavior that is not included in the model. In addition to that conclusion: female users are performing this unobserved behavior more than male consumers. Remarkably, female consumers have a negative relation to the stages of the theoretical framework, which implies a negative effect on remediation (lower remediation rate). Since the models estimate the opposite effect, the gap between stated and actual behavior must be substantial.

Our findings are illustrated in figure 39. Awareness, comprehension, intention and behavior are positively correlated and thus form a good backbone to understand how notifications lead to remediation. Some cleanup behavior is unobserved which results in the explanatory value of the variables awareness, intention and female. This latter covariate is remarkable since it has a negative relation to the stages, which indicates that women have cleaned up Mirai substantially more (intentionally or unintentionally) than they communicated during the interviews.

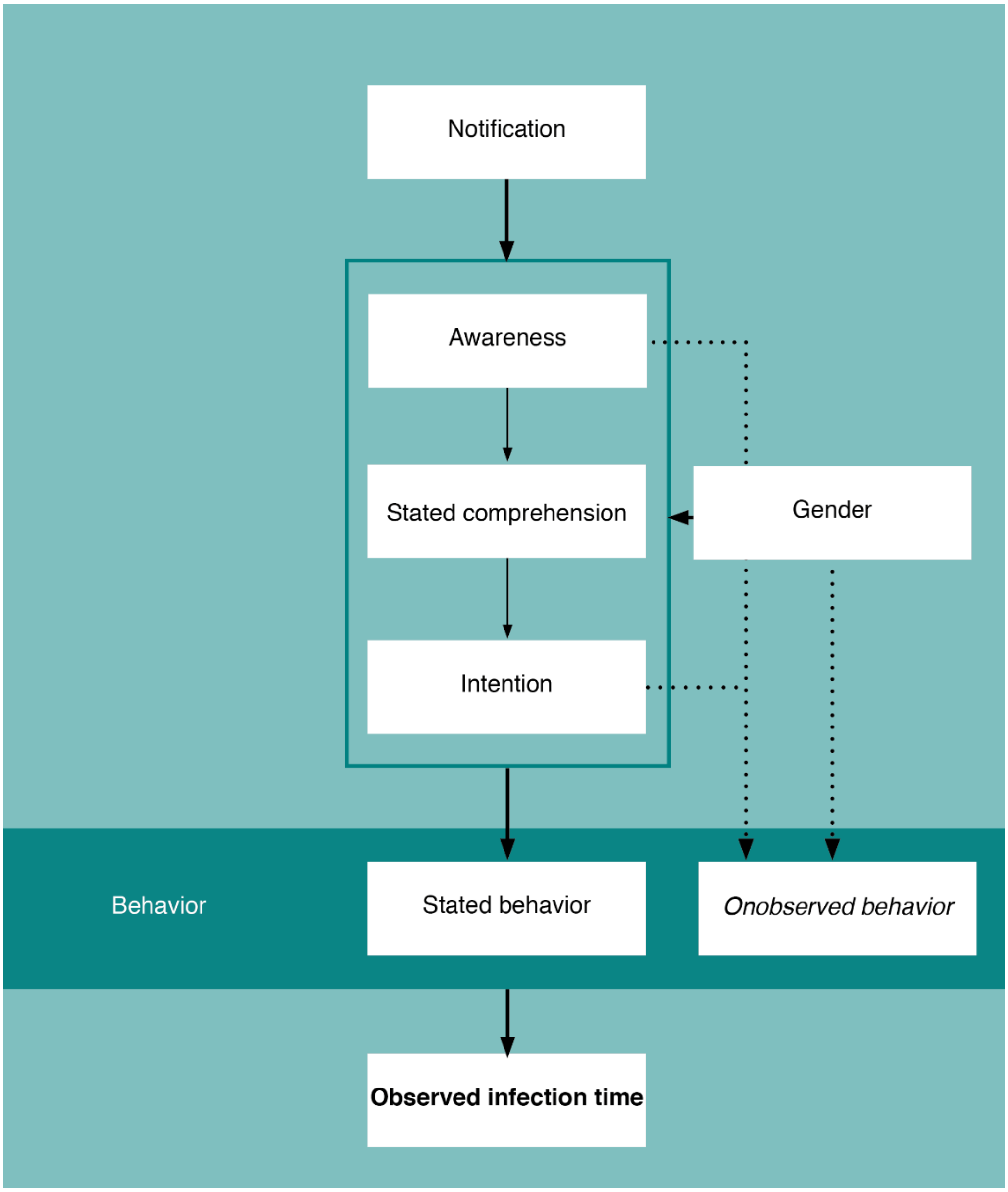


Figure 39 Theoretical framework adjusted to sub-conclusion on remediation drivers

11 Conclusions and discussion

11.1 Introduction

The objective of this research was to explore the role of IoT end users in Mirai-like bot remediation. To analyze this question, five sub-questions are formulated which have been answered in previous chapters. Section 11.2 will recap the main findings and provides the main conclusions. Section 11.3 elaborates on the implications of these conclusions for KPN and policy-making in general regarding Mirai remediation.

11.2 Main conclusions

To answer the question ‘what is the role of IoT device end users in Mirai-like bot remediation?’, we have conducted an eight-week experiment at the KPN Abuse Desk that notifies KPN and Telfort customers about abuse incidents. Mirai-infected consumers of these two markets have been randomly assigned (during staggered entry) to a walled garden, e-mail notification or control group. All 177 subjects within the experiment have been tracked for two weeks to estimate the infection time and are contacted afterward for interview purposes. Using different behavioral theories, we made a framework that serves as a backbone to understand how notifications influence remediation.

We conclude that male consumers were more exposed to Mirai infections during the experiment compared to female consumers. We can conclude the same for KPN consumers between the ages of 29 and 54 and Telfort consumers between the ages of 34 and 54. Consumers with this demographic background were relatively more in possession of a Mirai-infected device. One explanation is that consumers with this profile are more often in possession of an IoT device in general. Another explanation is that these consumers use their device differently (e.g., use it for more technically advanced applications, or use a device in a less secure manner), which increases the chance of Mirai-infection on these devices. In addition to these deviations, we observe a shifted age distribution of KPN Mirai-infected consumers in general. The mean age of this group is seven years younger than the mean age among all KPN Internet subscribers.

The majority of consumers do not follow the recommendations in the notification. In contrast, the number of actions that are performed that were not mentioned in the notifications is remarkably high. Consumers in the walled garden particularly take more drastic (but effective) measures such as disconnecting the device from the Internet, or discarding the infected device completely. We cannot point out the exact reason for this. One explanation is the lack of good comprehension. Since many consumers asked for additional help, we conclude that consumers appear don’t have a full understanding of how to tackle the problem. This may obstruct compliance. Another explanation is the lack of motivation to comply to all recommendations: consumers may not fully rely on the advice and prefer to solve the problem in their own way, or consumers may believe that they are not capable of performing the actions and thus take rigorous actions. In the control group, we find no clear explanation for remediation. One out of five consumers cleaned up Mirai unintentionally by updating outdated software; one consumer by the disconnection of devices. None of these consumers was aware of the fact that his/her device was infected with Mirai. Without the outbreak of a Mirai-variant exploiting a vulnerability of outdated software on home automation devices, the share of consumers who cleaned up their device would have been lower.

Both e-mail and walled garden notifications are effective in reaching the consumer, informing them, and encouraging them to take action. Consumers who are placed in a walled garden have as a primary incentive to get back Internet access while people who received an e-mail were motivated by the severity of the threat. There is a discrepancy between the stated comprehension of consumers and the observed comprehension. Although the majority of consumers stated that they understood the content of the notification completely, many consumers were not able to clean up their infected device without additional help. We also identified a gap between intention and behavior. Only 25% of the consumers who stated to be motivated to comply, succeeded in doing so completely. On the other hand, looking at compliance in loose sense (taking effective measures), the intention-behavior gap is smaller: 14% of the consumers did not manage to clean up their infected device.

In general, notified consumers are satisfied with KPN's effort to notify them about Mirai although this is not reflected in the communication with the Abuse Desk. A quarter of consumers placed in a walled garden are dissatisfied which can be explained by the disruptive nature of the measure. All customers who received an e-mail are satisfied with this service. The suggestions of consumers vary greatly but three recurring suggestions are a better information provision, availability of additional help and a better functioning walled garden notification process.

We find that gender and walled garden notifications have an influence on remediation. Consumers placed in a walled garden have a 29% to 85% shorter mean infection time. We can further specify the role of walled garden notifications: the covariates awareness (a consumer has received and read the content) and behavior (consumers has performed correct cleanup actions) explain remediation better than walled garden notifications alone. Since awareness of a notification has individual explanatory power - on top of the explanatory power of right behavior - we can conclude we do not observe all cleanup efforts. In other words: there is a discrepancy between *stated* behavior and *actual* behavior. Among notified consumers, intention explains remediation best. Since intention has more explanatory power than behavior, the same logic as before applies here: there is unobserved behavior which is influenced by the intention to comply.

Gender also influences remediation but the conservative estimates are not highly reliable: female consumers have a 1% to 87% shorter mean infection time than male consumers and subscribers of unknown gender. Following the logic as applied above, we cannot assume that Mirai-infected devices are cleaned up just because its user is female. Devices of female consumers are thus cleaned up more than the women in question stated. We conclude that age, consumer market, device type and customer satisfaction have no significant influence on remediation.

We provide four possible scenarios for the identified gap between the stated and unobserved behavior:

- Consumers forgot what clean up actions they performed or forgot to mention them during the interviews;

- Consumers clean up their device unintentionally;
- The device is cleaned up by someone else in the household without the consumer's knowledge;
- The consumer cleaned up the device after the interview: between the end of the experiment period and the end of the two-week additional observation.

Since the unexplained remediation rates are so high, one can argue that unobserved behavior is not able to explain that complete gap. That would imply that more than half of all consumers cleaned up their device while stating otherwise during the interviews. Since we believe this is unlikely, we must look for answers on the attacker side. We provide the following two unexplored explanations:

- Other malware takes over Mirai-infected devices. This study focuses on Mirai and only included abuse feeds on this malware. Therefore, a device may be compromised by other malware without our knowledge. That would explain why an IP address doesn't appear on our radar and the bot is wrongly considered as cleaned up.
- The majority of Mirai infections are detected when a bot is in a scanning phase. Conventional Mirai bots are scanning unless they get commands from the botnet herder. With the increasing number of Mirai-variants, bots may have evolved scanning behavior. They might for example only scan when commanded to, or have built-in behavior that determines that a bot only scans in the first hours of its life.

The findings are aligned with the theoretical framework and illustrated in figure 40.

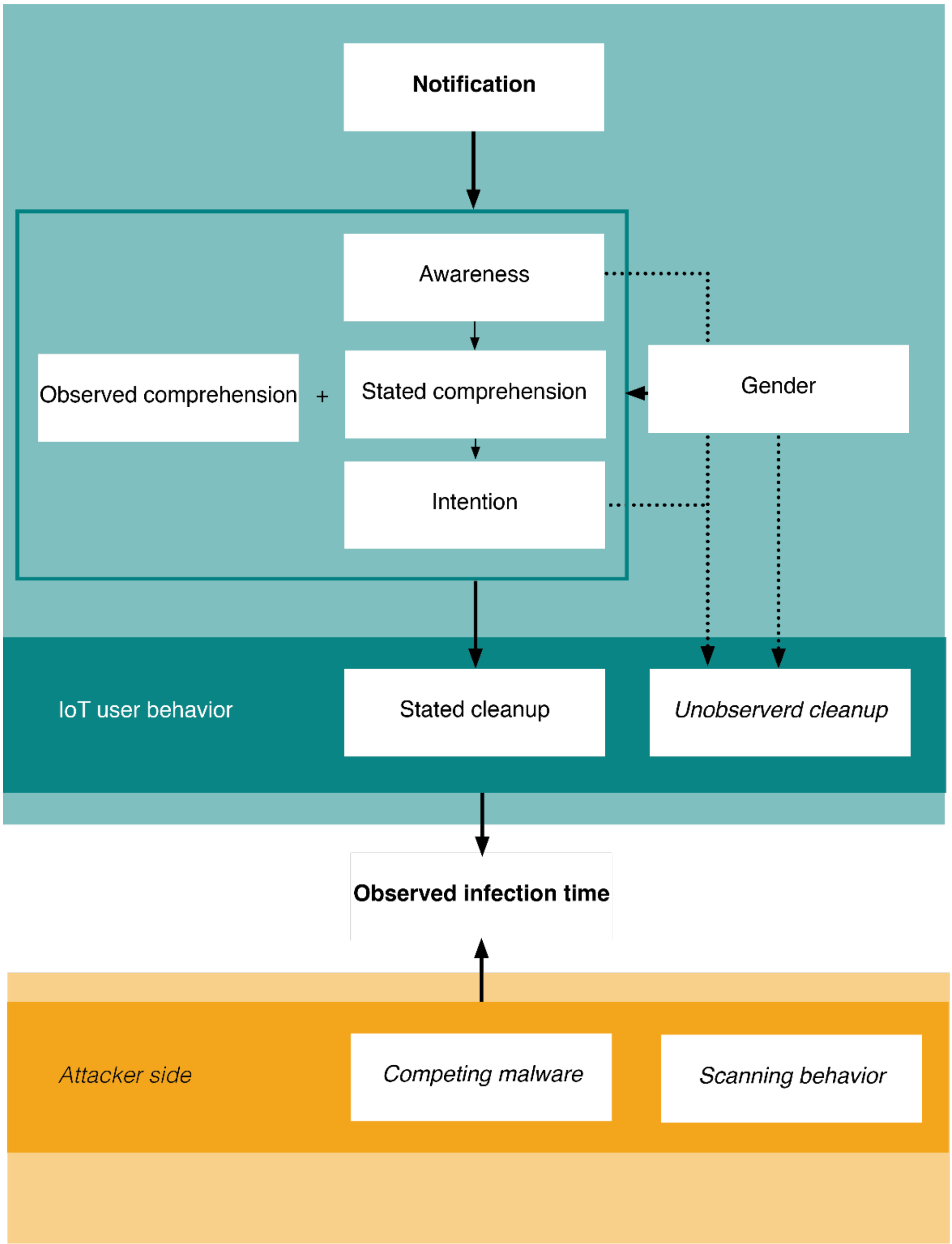


Figure 40 Theoretical framework adjusted to main conclusions

11.3 Implication of findings

11.3.1 Recommendations Abuse Desk

People wish to gain more information about the problem and how to resolve it. Due to the differences in technical knowledge, KPN could consider establishing an online information page with basic information, more elaboration on how to perform the steps and FAQs. Reference to such trusted website may also take away suspicion (since it enables consumers to fact-check).

A warning prior to placing consumers into a walled garden will prevent much confusion and improve customer satisfaction. The self-release option can be improved and needs to be mentioned in the e-mail notification to avoid unnecessary waiting and confusion among consumers, and saves extra work for the Abuse Desk. Although the contact form cannot be easily changed into a dynamic form, there can be made improvements in the static version. Currently, the introduction of the contact form refers to the 'problems on your computer/laptops', this can be changed to 'problems on your network'. In addition, the questions that are not applicable to all abuse incidents must explicitly state so. For example: the question about which virus scanner a consumer uses, can be complemented with the note that the consumer doesn't need to provide this information in case of a Mirai infection.

KPN could consider to better expose the activities of the Abuse Desk. Since this is a first-line service (direct contact with customers), consumers find it odd to have never heard of it before. This unfamiliarity, in combination with the notion 'abuse', creates distrust towards the notification. Awareness of the Abuse Desk's existence may improve consumer's co-operation and can also be marketed as a unique selling point in KPN's service. The content of notifications could be further personalized to prevent suspicion (e.g. inclusion of KPN account number). If the filing system of KPN permits, abuse incidents could even be included in the online logs of a consumer's account (MijnKPN) so that customers can verify the authenticity of the notification.

We would recommend the consideration of other means of communication that would complement the current notification practices. Sending an SMS is potentially very effective in A) making the consumer timely aware of the walled garden placement so there is no confusion about the cause of the disconnect and consumers are given the opportunity to solve the problem inside office hours, and B) increasing the trustworthiness of the other notification mechanisms because of the use of two channels.

Lastly, consumers will be helped greatly if the Abuse Desk and Help Desk are better integrated. A minimal requirement is that the Help Desk must be able to check the abuse incidents and notifications sent to a customer. Technically this requirement is met but in practice, Help Desk officers lack awareness and do not check automatically the Abuse Desk tickets. This can be solved by integrating the systems so that the Help Desk can monitor the Abuse Desk within the CRM. However, due to the complexity and multitude of Abuse Desk systems, this is easier said than done. In addition, the Help Desk must be better equipped with knowledge and access to tools so that can help customers who are placed in a walled garden. Although the Abuse Desk steers customers to only communicate per mail, customers often call the Help Desk in reality. This, in combination with unavailability outside office hours of the Abuse Desk, makes it worth to extend the capabilities of the Help Desk to help these customers.

11.3.2 Policy implications

Due to the exploratory nature of this research, the findings cannot be used for specific policy recommendations. Instead, the outcomes may be applied as background information to understand the problem at stake and to make better-informed policy decisions concerning IoT-botnet remediation. During the course of this study, the Dutch government has announced to increase its efforts to increase IoT security and mitigate IoT abuse (Ministerie van Economische Zaken en Klimaat, 2019). This study can be of added value within two parts of that roadmap: the development of an awareness campaign and the stimulation of IoT abuse mitigation by ISPs. These efforts can be enriched by the following takeaways:

Firstly, performing the correct remediation efforts is effective and walled garden notifications stimulate this behavior. This seems a trivial conclusion but it is important because it supports the decision to give ISPs a dominant role in IoT abuse mitigation. In addition to that, not all notified consumers succeed in performing effective remediation actions. This is because A) the remediation measures may differ per Mirai variant (as the Domoticz exploit illustrated) and B) consumers believe they have remediated but in reality have not (intention-behavior gap). Since we may expect more sophisticated and varying Mirai variants, there is no singular set of actions which will remediate all Mirai infections. This implies that instructions for Mirai remediation must be dynamic and case-specific.

Secondly, Mirai-victims cannot be captured in a few personas. The variety and increasing size of IoT devices cause a wide variety of victims: young, old, tech-savvy or not, etc. Informing potential victims may thus be challenging because different persons need different information.

The last suggestion is a follow-up on the idea of an online information page mentioned in the previous section. It would be helpful to develop such platform nation-wide in collaboration with all relevant stakeholders. In addition to basic information, this platform can be extended with dynamic updates on for example newly found exploits or detected outbreaks.

12 Reflection and future work

12.1 Introduction

This chapter reflects upon the research and conclusions. Section 12.2 discusses the research quality by addressing the limitations and their effect on the validity of the results. Section 12.3 provides ideas for future research.

12.2 Research quality

12.2.1 Limitations

One main limitation is the inaccurate measurement of infection time of Mirai bots. This limitation prevents us from making statements about exact remediation speed and rates. However, due to the modeling approach, we have been able to retrieve valuable information about the relative influence of factors.

Secondly, the experiment has not controlled for the Mirai variants. This is not possible due to the magnitude of variation and the absence of this information. Due to the Domoticz exploit, we tested whether that Mirai variant (and its inherent victims) plays a role. Although we could not control for Mirai-variant, we have included a dummy variable that distinguishes detections before and during the Domoticz-variant outbreak in the modeling steps. Since this variable is not estimated to have a significant influence, we conclude that this variant has no influence on observed remediation.

Thirdly, due to malfunctioning of the KPN mail server, we have not been able to obtain data from KPN customers who are only notified through e-mail. The e-mail treatment group in this study thus only exist of Telfort consumers which makes the results about e-mail notifications only valid for this population. In addition, the landing page of Telfort has been malfunctioning which prevented Telfort consumers from self-release. This is visible in the survival curves in chapter 6, but the influence is not big enough to obstruct proper analysis. However, one must realize the estimated infection time of this group would have been shorter within the first few days of infection if the landing page wouldn't malfunction.

Fourthly, we interviewed the control group after two weeks which can be regarded as a treatment. In theory, this might have decreased the number of censored observations, leading to an overestimated remediation speed and rate. However, the comparison of the survival curves of interviewed and non-interviewed consumers in the control group in section 7.2 shows that there is no significant difference in survival behavior (the non-interviewed consumers even perform better in terms of remediation).

Fifthly, we find that gender is of influence for remediation. However, since this has not been found before in other studies, we didn’t control for gender. As a consequence, women were relatively often assigned to the control group. In addition, the number of interviewed female consumers is low, making the estimates of the modeling steps 2 and 3 less reliable. We therefore only mention the conservative estimates of the first modeling steps in the conclusions.

Lastly, we miss one of the two daily feeds on sixteen days. This creates a blind spot in the detection of Mirai bots which may result in underestimated infection times. However, at least one feed was available each day so very active bots are likely to be detected.

12.2.2 Internal validity

We assume that all Mirai bots within the KPN and Telfort markets are included in the experiment. However, there is no possibility to cross-check these numbers. We can therefore not be completely certain that we have included the total Mirai-infected population.

The last limitation is the validity of the data obtained from the interviews. There are three things to take into account:

- Consumers may unknowingly give wrong answers. They may have forgotten what action they have performed or have identified the wrong device.
- Consumers may knowingly give wrong answers. Consumers may give answers they believe are desired because they have the feeling of being checked by KPN or they want to please the interviewer.
- Of all consumers in the experiment, 99 consumers are interviewed. Although this is a relatively high attendance, information about device type, actions, reasons for non-compliance and experience are not obtained from 78 consumers.

Due to the unobserved behavior we identified during modeling, we can conclude that what consumers have shared in the interviews does indeed not always match with reality. However, the data still provides a powerful first step in the exploration of consumers’ role in Mirai remediation.

12.2.3 External validity

We cannot infer the results to Mirai infection times in future populations due to the dynamic character of Mirai’s evolution and unpredictable behavior. In addition, users may also alter their behavior regarding IoT devices over time. However, we find no significant differences between the results of the same experiment in 2018, which implies that the behavior of Mirai and infected-device users have not substantially altered during the past year.

The research is culture and market-specific. KPN and Telfort have other target groups due to price differences and have other demographical compositions (Telfort consumers are younger as illustrated in section 5.3). Despite these differences, we have identified no significant difference between Telfort and KPN which is a promising result for the generalizability of the results to other Dutch ISPs. However, we cannot make the claim that our results also apply to other Dutch ISPs for two reasons: A) the number of Telfort consumers in the experiment is relatively low which causes an increased chance of Type II error, and B) more ISPs must be researched to support such claim.

In addition, the research only consists of Dutch ISPs. We cannot assume that IoT users in other countries have similar device types, user characteristics, and coping mechanisms.

12.3 Future work

The conclusions and limitations create respectively new knowledge gaps and room for improvement.

An important gap which needs to be covered is the lack of understanding of natural remediation as identified by Alterna (2018). This study has attempted to shine more light on this by exploring the role of infected users. We have concluded that the actions of consumers have a significant effect on remediation, but cannot explain all observed remediation. This means future research must focus more on the attacker side. We provide two scenarios which can both be explored:

The first scenario is that a Mirai-infected device is taken over by another malware. This would explain why Mirai is remediated without intervention from the user. This theory can be tested by analyzing abuse feeds: if an IP address is still detected while not being fingerprinted as Mirai, we know the theory is correct for that particular IP address. Preferably, only raw data from honeypots and darknet infrastructures are used since processed abuse feeds (as provided by Shadowserver) exclude detections that are not labeled. Even if we don't know the precise malware type, we still are interested in knowing whether a device is compromised by another malware than Mirai. Hajime is an example of an emerging malware that is known to compete with Mirai for IoT devices.

The second scenario is that some Mirai variants have different scanning behavior than we assumed. Remember: we detect the majority of Mirai-infected devices when they are in a scanning phase. Conventional Mirai bots scan the Internet in search of vulnerable devices unless they are given commands by the botnet herder. Hitherto we assume that Mirai bots are not constantly executing commands, and thus appear on our radar sooner or later. However, it is possible that certain variants have deviating built-in scanning behavior (e.g., only scan the first few hours of its lifecycle). That means we only detect a bot at the start of its life. Another explanation is that bots are given more commands than we expected so that a bot only scans for really short periods of time. This reduces the chance we detect a bot. Both explanations can be explored by collecting Mirai variants through honeypots and execute the malware code in a secured environment. However, this exploration will take much time due to the vast amount of Mirai variants.

Many consumers have not complied with the recommendations in the notification. Also, a number of interviewed consumers suggested a better information provision in the notification. These two findings indicate that the current notifications can be improved or that an external information source may be helpful. Future research could focus on how consumers react to different notification contents and what information is essential to reach more compliance.

We also observed a large intention-behavior gap and a gap between stated and actual behavior. These gaps can be further studied in a lab setting to observe what IoT users do in reality versus what they think/say they have done. In future research, it is recommended to control for gender since we conclude that gender influences the observed remediation. We have no explanation yet for the big difference in remediation between male and female IoT users. Possibly, women are more forgetful about their actions, or are unaware of other actions performed by others in a household. This needs further attention.

The last recommendation for future work is research in the behavior and demographics of IoT users in general. We observe that men and ‘younger’ (age 26-49) IoT users have relatively fallen more victim to a Mirai infection. However, we have no explanation yet for these findings. These groups may possess more IoT devices, deal differently with their device, or a combination of both.

References

Abuse Information Exchange. (n.d.). Home. Retrieved May 8, 2019, from https://www.abuseinformationexchange.nl/

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. https://doi.org/10.1016/0749-5978(91)90020-T

Almeida, V. A. F., & Goh, B. (2017). A Principles-Based Approach to Govern the IoT Ecosystem. *IEEE Internet Computing*, 78–81. https://doi.org/10.1109/MIC.2017.2911433

Altena, E. M. (2018). *Exploring effective notification mechanisms for infected iot devices*. TU Delft.

Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., … Kumar, D. (2017). Understanding the Mirai Botnet. *Proceedings of the 26th USENIX Security Symposium*. Retrieved from https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis

Aytes, K., & Connolly, T. (2004). Computer security and risky computing practices : A rational choice perspective. *Journal of Organizational and End User Computing*, *16*(2), 22–40.

Batalla, J. M., Mastorakis, G., Mavromoustakis, C., & Pallis, E. (Eds.). (2017). *Beyond the Internet of Things: Everything Interconnected*. https://doi.org/10.1007/978-3-319-50758-3

Bauer, J. M., & van Eeten, M. J. G. (2009). Cybersecurity: Stakeholder incentives, externalities, and policy options. *Telecommunications Policy*. https://doi.org/10.1016/j.telpol.2009.09.001

Bertino, E., & Islam, N. (2017). Botnets and Internet of Things Security. *Computer*, *50*(2), 76–79. https://doi.org/10.1109/MC.2017.62

Boss, S. R., Moody, G. D., Polak, P., Lowry, P. B., & Galletta, D. F. (2015). What Do Systems Users Have to Fear? Using Fear Appeals to Engender Threats and Fear that Motivate Protective Security Behaviors. *MIS Quarterly*, *39*(4), 837–864. https://doi.org/10.25300/misq/2015/39.4.5

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer*, *89*(3), 431–436. https://doi.org/10.1038/sj.bjc.6601119

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly*, *34*(3), 523–548.

Carretto, F. (2019). Domoticz 4.10577 Unauthenticated Remote Command Execution. Retrieved July 29, 2019, from Packet Storm website: https://packetstormsecurity.com/files/152678/Domoticz-4.10577-Unauthenticated-Remote-Command-Execution.html

Castillo-Montoya, M. (2016). Preparing for Interview Research: The Interview Protocol Refinement Framework. *The Qualitative Report*, *21*(5), 811–831. Retrieved from https://nsuworks.nova.edu/tqr/vol21/iss5/2

Çetin, Orçun, Altena, L., Gañán, C., & Eeten, M. Van. (2018). Let Me Out ! Evaluating the Effectiveness of Quarantining Compromised Users in Walled Gardens. In *Fourteenth Symposium on Usable Privacy and Security* (pp. 251–263). Retrieved from https://www.usenix.org/conference/soups2018/presentation/cetin

Çetin, Orçun, Gañán, C., Altena, L., Kasama, T., Inoue, D., Tamiya, K., … Van Eeten, M. (2019). Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai. *NDSS*. https://doi.org/10.14722/ndss.2019.23438

Çetin, Orcun, Gañán, C., Korczyski, M., & Van Eeten, M. (2017). Make Notifications Great Again: Learning How to Notify in the Age of Large-Scale Vulnerability Scanning. *16th Workshop on the Economics of Information Security (WEIS 2017)*, 1–23. Retrieved from http://mkorczynski.com/WEIS2017Cetin.pdf

Çetin, Orçun, Jhaveri, M. H., Gañán, C., van Eeten, M., & Moore, T. (2016). Understanding the role of sender reputation in abuse reporting and cleanup. *Journal of Cybersecurity*, *2*(1), 83–98. https://doi.org/10.1093/cybsec/tyw005

Cranor, L. F. (2008). A Framework for Reasoning About the Human in the Loop. *Proceedings of the 1st Conference on Usability, Psychology and Security*, 1–15. https://doi.org/10.1109/MSP.2010.198

Creswell, J. W. (2014). Research Design - Qualitative & Quantitative Approaches. In *Research Design: Qualitative and Quantitative Approaches* (4th ed.). California: SAGE Publications, Inc.

Dinev, T., & Hu, Q. (2007). The Centrality of Awareness in the Formation of User Behavioral Intention Toward Preventive Technologies in the Context of Voluntary Use. *Journal of the Association for Information Sysytems*, *8*(7), 386–408.

Donno, M. De, Dragoni, N., Giaretta, A., & Spognardi, A. (2018). DDoS-Capable IoT Malwares : Comparative Analysis and Mirai Investigation. *Security and Communication Networks*, 1–30.

Download. (2019). Retrieved July 29, 2019, from Domoticz website: https://www.domoticz.com/downloads/

Egelman, S., Cranor, L. F., & Hong, J. (2008). You ’ ve Been Warned : An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In ACM (Ed.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1065–1074). https://doi.org/10.1145/1357054.1357219

ENISA. (2017). *Baseline Security Recommendations for IoT in the context of Critical Information Infrastructures*. https://doi.org/10.2824/03228

ENISA. (2018). *IoT Security Standards Gap Analysis Mapping of existing standards against requirements on security and privacy in the area of IoT*. https://doi.org/10.2824/713380

ENISA. (2019). *ECSM 2018 Deployment Report*. https://doi.org/10.2824/0844

Essent. (2017). Smart home in Nederland. Retrieved March 26, 2019, from https://www.essent.nl/content/particulier/kennisbank/slim-huis/smart-home-nederland.html#

Evans, D. (2011). How the Next Evolution of the Internet Is Changing Everything. In *Cisco Internet Business Solutions Group (IBSG)*. https://doi.org/10.1109/IEEESTD.2007.373646

Exploit Database. (2019). Domoticz 4.10577. Retrieved July 6, 2019, from https://www.exploit-db.com/exploits/46773

Fagan, M., Khan, M. M. H., & Buck, R. (2015). A study of users’ experiences and beliefs about software update messages. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2015.04.075

Fagan, M., Maifi, M., & Khan, H. (2016). Why Do They Do What They Do?: A Study of What Motivates Users to (Not) Follow Computer Security Advice. *Twelth Symposium on Usable Privacy and*

Security, 59–75. Retrieved from https://www.usenix.org/conference/soups2016/technical-sessions/presentation/fagan

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*.

Felt, A., Ha, E., Egelman, S., & Haney, A. (2012). Android permissions: User attention, comprehension, and behavior. *Symposium on Usable Privacy and Security*. https://doi.org/10.1145/2335356.2335360

Forget, A., Pearman, S., Thomas, J., Acquisti, A., Christin, N., Cranor, L. F., … Telang, R. (2016). Do or Do Not, There Is No Try: User Engagement May Not Improve Security Outcomes. *Twelfth Symposium on Usable Privacy and Security*, 97–111. https://doi.org/oso/9780195183177.001.0001

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*. https://doi.org/10.4103/0974-7788.76794

Groenewegen, F. (2016). Ongekend grote DDoS-aanvallen, dit is nog maar het begin! - Fox-IT (NLD). Retrieved December 7, 2018, from https://www.fox-it.com/nl/insights/blogs/blog/ongekend-grote-ddos-aanvallen-is-nog-begin/

Gundu, T., & Flowerday, S. V. (2012). The enemy within: A behavioural intention model and an information security awareness process. *2012 Information Security for South Africa*, 1–8. https://doi.org/10.1109/ISSA.2012.6320437

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. https://doi.org/10.1080/10580530.2015.1117842

Heer, T., René, O. G., Loong, S., Sandeep, K., & Klaus, S. K. (2011). Security Challenges in the IP-based Internet of Things. *Wireless Personal Communication*, *61*(3), 527–542. https://doi.org/10.1007/s11277-011-0385-5

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: A framework for security policy compliance in organisations. *European Journal of Information Systems*. https://doi.org/10.1057/ejis.2009.6

Herley, C. (2009). So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. *NSPW '09 Proceedings of the New Security Paradigms Workshop*. https://doi.org/10.1145/1719030.1719050

Howe, A. E., Ray, I., Roberts, M., Urbanska, M., & Byrne, Z. (2012). The psychology of security for the home computer user. *Proceedings - IEEE Symposium on Security and Privacy*. https://doi.org/10.1109/SP.2012.23

Ifinedo, P. (2012a). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers and Security*. https://doi.org/10.1016/j.cose.2011.10.007

Ifinedo, P. (2012b). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers and Security*, *31*(1), 83–95. https://doi.org/10.1016/j.cose.2011.10.007

International Telecommunications Union. (2008). *ITU Botnet Mitigation Toolkit: Background information*. (January).

Jhaveri, M. H., Çetin, O., Gañán, C., Moore, T., & Eeten, M. Van. (2017). Abuse Reporting and the Fight Against Cybercrime. *ACM Computing Surveys*, *49*(4), 1–27. https://doi.org/10.1145/3003147

Klein, J., & Moeschberger, M. (2003). *Survival Analysis: Techniques for censored and truncated data* (2nd ed.). New York: Springer.

Kolias, C., Kambourakis, G., Stavrou, A., & Voas, J. (2017). DDoS in the IoT : Mirai and other botnets. *Computer*, 80–84. https://doi.org/10.1109/MC.2017.201

KPN. (n.d.). Abuse. Retrieved December 19, 2018, from https://www.kpn.com/service/internet/veilig-internetten/abuse.htm

Krol, K., Moroz, M., & Sasse, M. A. (2012). *Don't Work – Can't Work : Why It's Time to Rethink Security Warnings*.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science*, *9*, 181–211. https://doi.org/10.28945/479

Li, F., Bailey, M., Durumeric, Z., Czyz, J., Karami, M., Mccoy, D., … Paxson, V. (2016). You've Got Vulnerability: Exploring Effective Vulnerability Notifications. *Proceedings of the 25th USENIX Security Symposium*, 1033–1050. Retrieved from https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/li

Liang, H., & Xue, Y. (2018). Understanding Security Behaviors in Personal Computer Usage: A Threat Avoidance Perspective. *Journal of the Association for Information Systems*. https://doi.org/10.17705/1jais.00232

Lifelines. (2019a). Retrieved June 22, 2019, from https://lifelines.readthedocs.io/en/latest/index.html

Lifelines. (2019b). More examples and recipes. Retrieved June 20, 2019, from https://lifelines.readthedocs.io/en/latest/Examples.html?highlight=log-rank

Lindsey, J. K., Gill, R., Brian, D., Ross, S., Silverman, B. W., & Stein, M. (2004). *Statistical Analysis of Stochastic Processes in Time*. Cambridge University Press.

Livingood, J., Mody, N., & O'Reirdan, M. (2012). Recommendations for the Remediation of Bots in ISP Networks (RFC 6561). *Internet Eng. Task Force*, 1–29.

Ministerie van Economische Zaken en Klimaat. (2019). *Kamerbrief Voortgang roadmap digitaal veilige hard- en software* (pp. 1–10). pp. 1–10.

Mocrii, D., Chen, Y., & Musilek, P. (2018). IoT-based smart homes: A review of system architecture, software, communications, privacy and security. *Internet of Things*, *1–2*, 81–98. https://doi.org/10.1016/j.iot.2018.08.009

Neal, A. (2015). What's more general than a whole population? *Emerging Themes in Epidemiology*, *12*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4549103/

Ng, B. Y., Kankanhalli, A., & Xu, Y. (Calvin). (2009). Studying users' computer security behavior: A health belief perspective. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2008.11.010

O'Quigley, J. (2018). *Proportional Hazards Regression*. Retrieved from https://ebookcentral-proquest-com.tudelft.idm.oclc.org/lib/delft/detail.action?docID=337200

Online Trust Alliance. (2012). *Combatting Botnets Through User Notification Across the Ecosystem: a view of emerging practices*.

Raad Cyber Security. (2017). *Advies inzake de cybersecurity van het Internet of Things (IoT)*. *3*.

Rajab, M., & Eydgahi, A. (2019). Evaluating the explanatory power of theoretical frameworks on intention to comply with information security policies in higher education. *Computers & Security*, *80*, 211–223. https://doi.org/https://doi.org/10.1016/j.cose.2018.09.016

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *Journal of Psychology*, *91*, 93–114.

Safa, N. S., Sookhak, M., Von Solms, R., Furnell, S., Ghani, N. A., & Herawan, T. (2015). Information security conscious care behaviour formation in organizations. *Computers & Security*, *53*, 65–78. https://doi.org/https://doi.org/10.1016/j.cose.2015.05.012

Saikia, R., & Barman, M. P. (2017). Comparing Accelerated Failure Time Models with Its Specific Distributions in the Analysis of Esophagus Cancer Patients Data. *International Journal of Computational and Applied Mathematics*, *12*(2), 411–424.

Shadowserver. (n.d.-a). Data collection. Retrieved June 22, 2019, from https://www.shadowserver.org/what-we-do/data-collection/

Shadowserver. (n.d.-b). Drone/Botnet-Drone Report. Retrieved May 8, 2019, from https://www.shadowserver.org/what-we-do/network-reporting/drone-botnet-drone-report/

Silverio-Fernández, M., Renukappa, S., & Suresh, S. (2018). What is a smart device? - a conceptualisation within the paradigm of the internet of things. *Visualization in Engineering*, *6*(1), 3. https://doi.org/10.1186/s40327-018-0063-8

Siponen, M., Adam Mahmood, M., & Pahnila, S. (2014). Employees’ adherence to information security policies: An exploratory field study. *Information and Management*. https://doi.org/10.1016/j.im.2013.08.006

Sommestad, T., & Hallberg, J. (2013). *A Review of the Theory of Planned Behaviour in the Context of Information Security Policy Compliance*. https://doi.org/10.1007/978-3-642-39218-4_20

Statista. (2019). Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions). Retrieved March 27, 2019, from https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

Stock, B., Pellegrino, G., Li, F., Backes, M., & Rossow, C. (2018). Didn’t You Hear Me? - Towards More Successful Web Vulnerability Notifications. *Proceedings 2018 Network and Distributed System Security Symposium*, (February). https://doi.org/10.14722/ndss.2018.23171

Surf.nl. (n.d.). The SURF cooperative. Retrieved June 22, 2019, from https://www.surf.nl/en/about-surf/the-surf-cooperative

Surfnet.nl. (2018). Thunderlab home. Retrieved June 22, 2019, from https://wiki.surfnet.nl/display/THUN/Thunderlab+Home

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. https://doi.org/https://doi.org/10.1016/j.cose.2017.07.003

Torten, R., Reaiche, C., & Boyle, S. (2018). The impact of security awarness on information technology professionals’ behavior. *Computers & Security*, *79*, 68–79. https://doi.org/https://doi.org/10.1016/j.cose.2018.08.007

van Eeten, M., & Bauer, J. M. (2009). Emerging Threats to Internet Security: Incentives, Externalities and Policy Implications. *Ssrn*, *17*(4). https://doi.org/10.1111/j.1468-5973.2009.00592.x

van Eeten, M. J., & Bauer, J. M. (2008). Economics of malware: security decisions, incentives and externalities. In *Information and Communication Technologies*.

Vance, A., Siponen, M., & Pahnila, S. (2012). Motivating IS security compliance: Insights from Habit and Protection Motivation Theory. *Information and Management*, *49*(3–4), 190–198. https://doi.org/10.1016/j.im.2012.04.002

Vasek, M., & Moore, T. (2012). Do Malware Reports Expedite Cleanup? An Experimental Study. *Proceedings of the 5th Workshop on Cyber Security Experimentation and Test*, 1–8. https://doi.org/10.1016/j.egypro.2011.02.120

Vlajic, N., & Zhou, D. (2018). IoT as a land of opportunity for DDoS hackers. *Computer*, 26–34. https://doi.org/10.1109/MC.2018.3011046

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Review. *Management Information Systems Quarterly*, *26*(2), xiii–xxiii. https://doi.org/10.2307/4132319

Wogalter, M. S. (2006). *Handbook of warnings*. New Jersey: Lawrence Erlbaum Associates, Inc.

Zare, A., Hosseini, M., Mahmoodi, M., Mohammad, K., Zeraati, H., & Holakouie Naieni, K. (2015). A Comparison between Accelerated Failure-time and Cox Proportional Hazard Models in Analyzing the Survival of Gastric Cancer Patients. *Iranian Journal of Public Health*, *44*(8), 1095–1102. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26587473%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4645729

Appendices

A Literature search

The literature search is conducted through the framework as proposed by Webster and Watson (2002) and Levy and Ellis (2006). They propose the following three steps in identifying relevant literature:

1. Keyword search: the initial step using key words in scholarly databases and leading journals;
2. Backward search: reviewing citations of (relevant) articles;
3. Forward search: use academic search engines to find articles that has cited the (relevant) articles.

Details on the **keyword search** are presented in table 11. Three searches are executed in two journals (Computer & Security and Journal of Cybersecurity) and one database (IEEE Explore). The search terms were:

- security AND (perception* OR behaviour*);
- security AND (IoT OR 'Internet of Things');
- (malware OR 'malicious software*' OR 'botnet') AND (warning* OR notification*).

The Journal of Cybersecurity sometimes gave very few hits thus two search terms were altered for this journal. After these searches, the following steps were taken:

1. All hits were sorted by relevance by the search engine;
2. Depending on the actual relevance, 50 or 100 hits were studied (less if there were not many hits);
3. The relevant hits were added a list and structured

Table 12 indicates the outcomes of each search in terms of hits, studied articles (abstract), the number of relevant articles (added to the list) and how many of those were new to the list.

#	Journal/source	Search term	Hits/studied/relevant/(new)	Date
1.1	Computer & Security	security AND (perception* OR behaviour*)	1487/50/19	28/02/19
1.2		security AND (IoT OR 'Internet of Things')	611/50/10/4	28/02/19
1.3		malware OR 'malicious software*' OR 'botnet') AND (warning* OR notification*)	377/100/2/2	01/03/19
2.1	IEEE Explore	security AND (perception* OR behaviour*)	3575/50/11	28/02/19
2.2		security AND (IoT OR 'Internet of Things')	7354/50/16	28/02/19
2.3		(malware OR 'malicious software*' OR 'botnet') AND (warning* OR notification*)	95/50/2/2	01/03/19

3.1	Journal of Cybersecurity	security AND behaviour	45/45/5/5	28/02/19
3.2		*(security AND IoT OR 'Internet of things')*	3/3/0/0	28/02/19
3.3		*(malware AND notification)*	11/11/3/1	01/03/19

Table 12 Keyword search

For the **back- and forward search**, six most relevant articles from the list were reviewed. Table 13 contains similar information as table 12, only the 'number of hits' is replaced by the number of citations. The forward searches were conducted through the academic search engines Semantic Scholar and researchgate.net.

#	**Article**	**Backward search** **Citations/studied/ relevant/(new)**	**Forward Citations/studied/ relevant/(new)**	**Date**
1	Çetin, O., Jhaveri, M. H., Gañán, C., van Eeten, M., & Moore, T. (2016)	16/16/2/2	18/18/7/7	04/03/19
2	Thompson, N., McGill, T. J., & Wang, X. (2017)	96/96/23/21	6/6/2/1	04/03/19
3	Torten, R., Reaiche, C., & Boyle, S. (2018)	44/44/12/6	0/0/0/0	04/03/19
4	Pijpker, J., & Vranken, H. (2016)	25/25/9/9	2/2/0/0	04/03/19
5	Çetin, O., Altena, L., Gañán, C., & Eeten, M. Van. (2018)	31/31/12/6	1/1/1/1	04/03/19
6	Forget, A., Pearman, S., Thomas, J., Acquisti, A., Christin, N., Cranor, L. F., ... Telang, R. (2016)	40/40/8/7	0/0/0/0	04/03/19

Table 13 Back- and forward search

B Consulted literature

	Reference	*Aim*	*Relevant methodology*	*Relevant results/conclusions*	*Knowledge gaps*
Efficacy of (IoT abuse) notifications	Don't Work – Can't Work: Why It's Time to Rethink Security Warnings (Krol, Moroz, & Sasse, 2012)	One of the first studies into security warning effectiveness (download warning)	• Use of folk models (Wash) during interviews	• Security warnings are largely ineffective • Content does not matter. • Those with a lack of computer experience perform better. • Participants rely on their own judgment, rather than a security warning.	
	Do Malware Reports Expedite Cleanup? An Experimental Study (Vasek & Moore, 2012)	This paper describes assesses 'whether sending [abuse] reports to affected parties makes a measurable difference in cleaning up malware.'	• A relevant study design	• 'including details describing the compromise is essential [..] – sending reports with minimal descriptions of the malware is ineffective' • Sending multiple notices does not make an impact (compared to one notice)	• Impact of sender reputation (see Çetin, Jhaveri, Gañán, van Eeten, & Moore, 2016) • Reasons for re-infections are (see Orçun Çetin et al., 2019)
	Combatting Botnets Through User Notification Across the Ecosystem: a view of emerging practices (Online Trust Alliance, 2012)	This paper present the botnet notification best practices from a multi-actor perspective		• Tips to Improve the Delivery and Design of User Notifications • Preliminary List of Best Practices	

Do or Do Not, There Is No Try: User Engagement May Not Improve Security Outcomes (Forget et al., 2016)	‘This paper presents a qualitative study comparing users’ attitudes, [behaviors], and understanding of computer security to the actual states of their computers.’		• ‘User engagement alone may not be predictive of computer security.’ • Need for ‘concise, precise, simple, and easy-to-perform security instruction, [..] once applied, will remain effective without any user effort’	• ‘A need for a more critical evaluation of the content, presentation, and functionality of security interventions’ • More research needed into security interventions tailored to users with different levels of expertise
You've Got Vulnerability: Exploring Effective Vulnerability Notifications (Li et al., 2016)	This paper illuminates which aspects of vulnerability notifications (to non-end users) have the greatest impact on efficacy.		• Notifications improved remediation behavior (additional 11%) but most organizations did not patch their host	• No understanding of why these results are so modest • [no focus on *abuse* notification or *IoT abuse*]
Understanding the role of sender reputation in abuse reporting and cleanup (Orçun Çetin et al., 2016)	This study researches whether sender reputation is a driver of response to abuse notification	• A relevant research design	• ‘detailed abuse reports significantly increase cleanup rates.’ • There is ‘no evidence that sender reputation improves cleanup’	• ‘Remarkably little research has been undertaken into what factors drive the chances of a recipient acting upon an abuse report’
Make Notifications Great Again: Learning How to Notify in the Age of Large-Scale Vulnerability Scanning	This paper analyzed ‘the aspects and factors that drive vulnerability remediation rates and	• Use of survival probabilities to visualize remediation rate	• While notifications did lead to more remediation than in the control groups, the overall remediation rates were low.’	• The incentive structure for remediation are not well understood

(Orcun Çetin, Gañán, Korczyski, & Van Eeten, 2017)	how recipients feel about various types of notifications.’			
Didn't You Hear Me? - Towards More Successful Web Vulnerability Notifications (Stock, Pellegrino, Li, Backes, & Rossow, 2018)	This paper analyses the technical and human aspects that affect the success of vulnerability notifications.	• Variable: ‘aware-to-fix rate represents the chance that an issue is fixed after the report was viewed’	• The content of a notification is important in convincing operators to take action (discrepancy between problem awareness and addressing it) • E-mail as a communication medium suffers from several shortcomings but other channels do not justify their significant financial costs and time overheads.	• Incentives for remediation not well understood
Let Me Out ! Evaluating the Effectiveness of Quarantining Compromised Users in Walled Gardens (Orçun Çetin et al., 2018)	This paper user behavior and remediation effectiveness of walled gardens as a notification mechanism	• A relevant study design	• IoT malware remediation methods will differ from traditional clean-up strategy. • ‘Substantial support for the effectiveness of walled gardens’ for ISPs in the fight against botnets • Walled gardens may create a prisoners’ dilemma in ISP’s remediation efforts	
Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer	The first ‘empirical study of IoT malware cleanup [..] – more specifically, of removing Mirai	• A relevant study design	• E-mail only notifications did not have impact compared with the control group • High natural remediation rate of 58-74 %	• Not well understood what customers did after receiving a notification

	Efforts to Remove Mirai (Orçun Çetin et al., 2019)	infections in the network of a medium-sized ISP.’		• A very low reinfection rate • Walled gardens are effective but are not a large-scale solution	• A discrepancy between lab results and empirical results of reinfection
Notification processing	Handbook of warnings (Wogalter, 2006)	This book describes warning design standards and guidelines.	• The Communication-Human Information Processing Model (C-HIP) for structuring warning research		
	You’ve Been Warned : An Empirical Study of the Effectiveness of Web Browser Phishing Warnings (Egelman, Cranor, & Hong, 2008)	This study compared the effectiveness of active and passive phishing warnings by analyzing them using the C-HIP model.	• A warning analysis methodology: the C-HIP model.	• Active warnings (disturbing in user’s activity) are more effective than passive.	
	A Framework for Reasoning About the Human in the Loop (Cranor, 2008)	This article proposes a framework, largely based on the C-HIP model, to explain potential reasons for human failure in a cybersecurity context.		• Security actions are often to be performed by non-experts who are instructed in what to do (warnings, notification, etc.) Therefore, failure of such action can also be seen as a problem of incomplete communication.	
	Android permissions: User attention,	This study examines whether Android permission system is	• ‘Each step [within C-HIP model] is critical: a failure of usability at any step will	• ‘Most users fail to pass the attention and comprehension steps’	• Hypothesis that different users have different types of

comprehension, and behaviour (Felt, Ha, Egelman, & Haney, 2012)	effective at warning users.	render all subsequent steps irrelevant.’		privacy and security concerns and that addressing those in a warning will make the warning more effective
A study of users' experiences and beliefs about software update messages (Fagan, Khan, & Buck, 2015)	This study explores the relation between beliefs in different software updates and the effectiveness of those.	• Survey questions based on C-HIP	• Most users are annoyed by software warning and update messages, which affects the attitude/belief stage in the C-HIP model which causes more non-compliance.	
Computer security and risky computing practices: A rational choice perspective (Aytes & Connolly, 2004)	Why people who are aware of of the risks, still expose insecure behavior, as an outcome of a boundedly-rational choice process	• Based on conditions founded in TRA and TAM	• People don’t make sensible action decisions and therefore it is unlikely that additional information on risks improves behavior	• Further understanding of factors that influence decision process is needed.
Economics of malware: security decisions, incentives and externalities (M. J. van Eeten & Bauer, 2008)	This working paper reports on qualitative empirical research into the incentives of market players when dealing with malware.	• ‘Many of the problems of information security can be explained more clearly and convincingly using the language of microeconomics’	• The development of a 'culture of security' is very sensitive to economic incentive structures • Overview of externalities	
Emerging Threats to Internet Security: Incentives, Externalities and Policy Implications	This study explains the causes of the rise in botnets by the incentive structure of market players.	• A multi-actor perspective • RCT as explanation for both behavior as well as the aggregate outcome to society	• Machine owners have little incentive to remediate a botnet • End users’ behavior enables the growth of botnets, which impose costs on every other actor in the network.	

(M. van Eeten & Bauer, 2009)			• Benefits in cost/benefit trade-offs are rather 'potential costs to society of attacks that have not yet occurred.'	
Cybersecurity: Stakeholder incentives, externalities, and policy options (Bauer & van Eeten, 2009)	'The paper develops a framework for studying the co-evolution of the markets for cybercrime and cybersecurity.'		• 'Market and non-market relations in the information infrastructure generate many security-enhancing incentives. However, pervasive externalities remain that can only be corrected by voluntary or government-led collective measures.'	
So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users (Herley, 2009)	This article argues that 'users' rejection of the security advice they receive is entirely rational from an economic perspective.'	• Working security advice when: d(benefit) > d(costs)	• Benefits are overestimated whole costs of user effort is often ignored. • Users are rational, they 'only' need a better understanding of the harms they face	
Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness (Bulgurcu et al., 2010)	This research identifies rationality based factors that drive an employee to comply with information security policy of an organization.	• Combines rational choice theory with theory of planned behavior • Use of structural model testing (PLS approach)	• 'along with normative belief and self-efficacy, an employee's attitude toward compliance determines intention to comply. • We posit that an employee's attitude is influenced by the benefit of compliance, cost of compliance, and cost of noncompliance'	

Why Do They Do What They Do?: A Study of What Motivates Users to (Not) Follow Computer Security Advice (Fagan, Maifi, & Khan, 2016)	This paper investigates user motivation to follow-up on computer security advice.	• Cost/benefit framework can be used to investigate motivation of users to follow computer security advice.	• Choices in the experiment provided more perceived benefit than costs (compliant and in noncompliant decisions) • 'Social considerations are trumped by individualized rationales.'	There is a gap between perceived and actual costs and benefits.
Abuse Reporting and the Fight Against Cybercrime (Jhaveri et al., 2017)	This paper presents a model of the abuse reporting infrastructure to improve the understanding of voluntary actions against cyber crime		• 'Because no single entity is responsible for reporting, maintaining, and acting on abuse data, incentives determine why participants take action'	• 'For the immediate future, it seems more promising to increase the effectiveness of the abuse reporting infrastructure within the existing incentive structure.'

Table 14 Consulted literature

C Behavioral models

Abbr.	*Theory/model*	*Explains:*	*Relevant? (+ or -)*	*Named in:*	*Developed by:*
ARI	Affect-Reason-Involvement model	How users can be convinced to comply with a message in three ways (rational appeal, emotional appeal or both)	- difficult variables to make tangible - underrepresented in security studies	(Fagan et al., 2015)	Buck, 1994
BH	Health belief model	Healthcare behavior based on expectancy-value principles (Ng, Kankanhalli, & Xu, 2009)	+ has been widely applied to many domains (Ng et al., 2009) - similar construct as PMT but PMT is more applicable in this domain as argued by (Hanus & Wu, 2016)	(Ng et al., 2009) (Hanus & Wu, 2016)	Rosenstock, 1966
C-HIP	Communication-Human Information Processing model	How communication to an individual triggers his/her behavior and to identify reasons why notifications may be ineffective	+ framework to understand notification processing by end users + is used as a backbone in four earlier studies to systematically analyse failure of the desired behavior	See four articles under 'Notification processing' in table 14	Wogalter, 2006
CET	Cognitive Evaluation Theory	detrimental effects of rewards on intrinsic motivation, especially when rewards were tangible (Siponen, Adam Mahmood, & Pahnila, 2014, p.219)	- rewards not relevant in this research	(Siponen et al., 2014)	
FM	Folk models	Conceptualizations of home computer security threats (Forget et al., 2016)	+ focus particular on botnets - not a comprehensive or validated theory	(Krol et al., 2012) (Forget et al., 2016) (Orçun Çetin et al., 2019)	Wash, 2010
GDT	General Deterrence Theory	' the effect of deterrent factors on security policy compliance.' (Herath & Rao, 2009, p.109)	- The theory proposes that non-compliance can be deterred with severe punishment. Punishment is not desired in this research. - Rajab & Eydgahi (2019) find little support for GDT to explain variance	(Herath & Rao, 2009) (Rajab & Eydgahi, 2019)	Williams & Hawkins, 1986

			in information security policy compliance		
GEMS	Generic-Error Modeling System	'human security failures' and distinguishes three types of human errors (Cranor, 2008)	+ addresses the gap between intention and actual behavior which is not covered by other models - ignores many aspects of behaviour	(Cranor, 2008)	Reason, 1990
KAP	Theory of Knowledge, Attitude and Practice	effectiveness of training in terms of change in attitudes and behavior (Torten, Reaiche, & Boyle, 2018)	- no training in notification processes	(Torten et al., 2018)	
NT	Neutralization theory	Behavior as an outcome of a 'rationale to justify actions and neutralize guilt' (Torten et al., 2018, p.69)	- It studies behaviour after action (post hoc) and does not help explaining how behaviour can be modified	(Torten et al., 2018)	Matza & Sykes, 1964
PMT	Protection Motivation Theory	Has evolved greatly from the theory of fear appeal to model that is used to explain risky behavior.	+ many studies show the explanatory value of the model + based on TPB and TRA (Boss, Moody, Polak, Lowry, & Galletta, 2015)	21 studies using PMT for explaining cybersecurity behaviour	Rogers, 1975 (Hanus & Wu, 2016 modified the model for security behavior)
RCT	Rational Choice Theory	Behavior as the outcome of a cost-benefit trade-off. Often combined or complemented with another model (TPB, TRA, TAM)	+ is complementary to other models + powerful rationale to explain the rise of botnets (as the outcome of incentive structure) - actors are not fully rational (no complete information on benefits and costs, perception and beliefs play a major role)	See eight articles under 'IoT abuse remediation' in table 14	A neo-classical economic approach

TAM	Technology Acceptance Model	'Attitude and its antecedents (behavioral beliefs)' as the outcome of 'objective information concerning information technologies and their design' (Bulgurcu et al., 2010)	+ IoT can be considered a new technology and therefore TAM may explain the influence of its adaption on behavior - TAM does not explain user behavior well of protective technologies (Dinev & Hu, 2007) - Same foundation as TPB (both an extension of TRA) (Dinev & Hu, 2007) but TPB more applicable	(Ng et al., 2009) (Mocrii, Chen, & Musilek, 2018) (Howe et al., 2012) (Dinev & Hu, 2007)	Davis, 1989
TPB	Theory of Planned Behavior	Explores 'intentions prior to actions, which is driven by the values of the individuals to behave' (Torten et al., 2018, p.69)	+ more general version of the PMT (Thompson et al. 2017) + multiple times used in studies into information security policy compliance - requires a large qualitative study to interpret behavior (Torten et al. 2018)	(Thompson, McGill, & Wang, 2017) (Ifinedo, 2012b) (Bulgurcu et al., 2010) (Rajab & Eydgahi, 2019)	Ajzen, 1985
TRA	Theory of Reasoned Action	TRA predicts behavior by a person's intention to take actions, which is influenced by a person's attitude and subjective norms	+ intention is not similar to actual action (intention-behavior gap) - the predecessor of TPB (Bulgurcu et al., 2010; Sommestad & Hallberg, 2013)	(Gundu & Flowerday, 2012) (Sommestad & Hallberg, 2013) (Bulgurcu et al., 2010)	Fishbein and Ajzen 1975
TTAT	Technology Threat Avoidance Theory	'why and how individuals avoid IT threats in voluntary settings' (Liang & Xue, 2018)	- quite similar as PMT as constructed by Hanus and Wu (2016) but only one study identified that uses this model	(Liang & Xue, 2018)	Liang & Xue, 2018

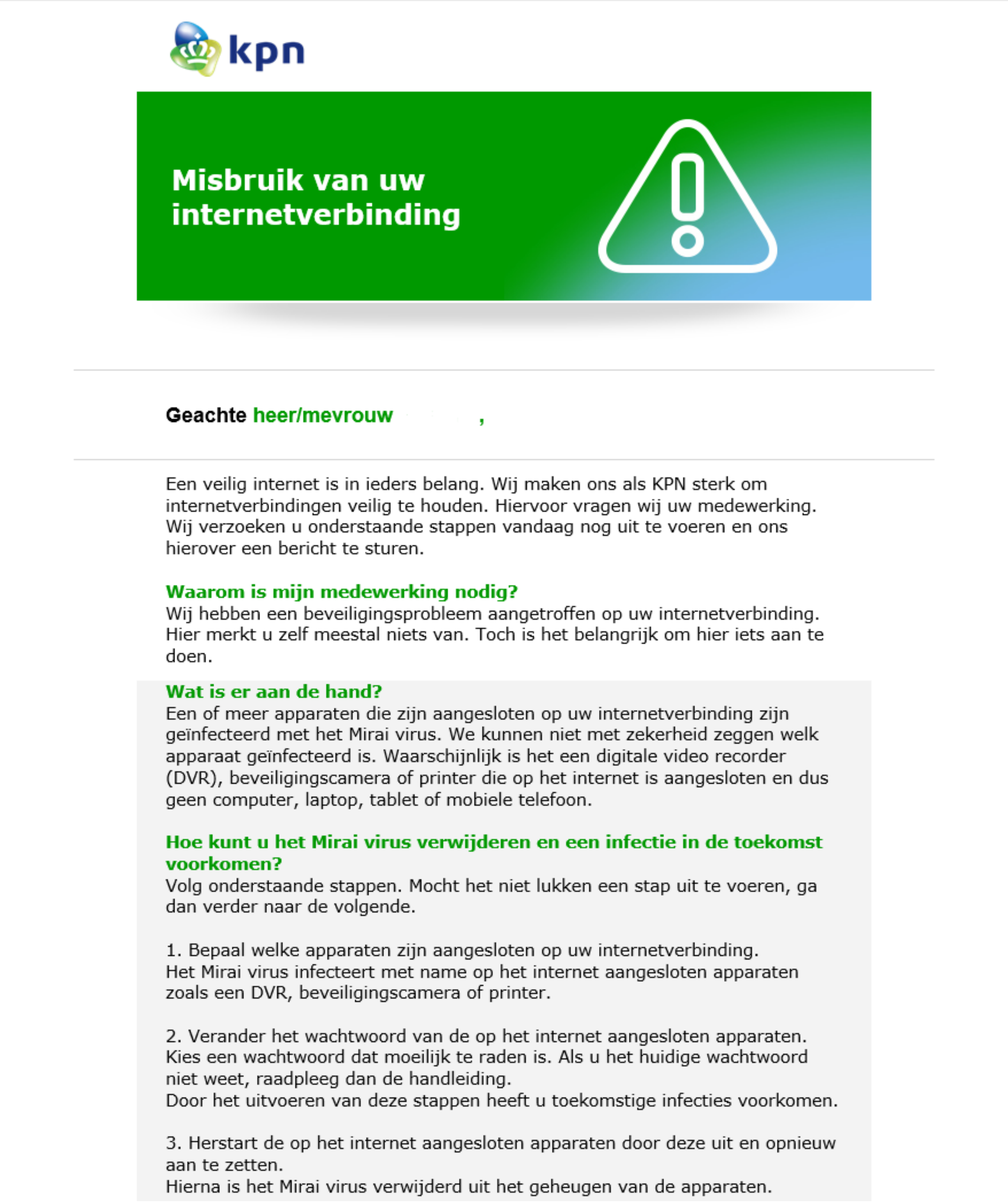
Table 15 Behavioral models

D Abuse Team procedures

D.1 & D.2

CONFIDENTIAL

D.3 Notification e-mails, landing page and contact form



Misbruik van uw internetverbinding

Geachte heer/mevrouw ,

Een veilig internet is in ieders belang. Wij maken ons als KPN sterk om internetverbindingen veilig te houden. Hiervoor vragen wij uw medewerking. Wij verzoeken u onderstaande stappen vandaag nog uit te voeren en ons hierover een bericht te sturen.

Waarom is mijn medewerking nodig?
Wij hebben een beveiligingsprobleem aangetroffen op uw internetverbinding. Hier merkt u zelf meestal niets van. Toch is het belangrijk om hier iets aan te doen.

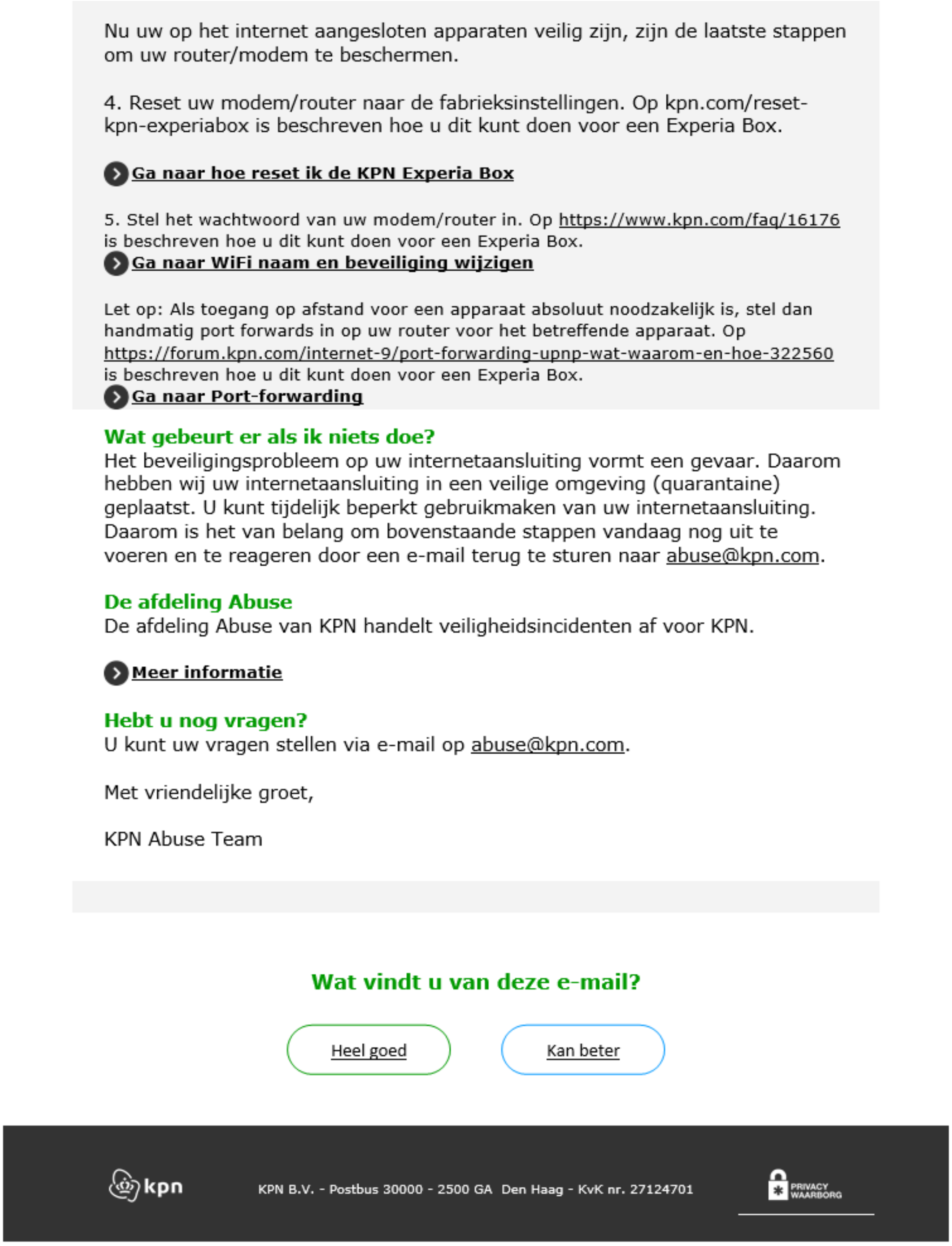
Wat is er aan de hand?
Een of meer apparaten die zijn aangesloten op uw internetverbinding zijn geïnfecteerd met het Mirai virus. We kunnen niet met zekerheid zeggen welk apparaat geïnfecteerd is. Waarschijnlijk is het een digitale video recorder (DVR), beveiligingscamera of printer die op het internet is aangesloten en dus geen computer, laptop, tablet of mobiele telefoon.

Hoe kunt u het Mirai virus verwijderen en een infectie in de toekomst voorkomen?
Volg onderstaande stappen. Mocht het niet lukken een stap uit te voeren, ga dan verder naar de volgende.

1. Bepaal welke apparaten zijn aangesloten op uw internetverbinding.
Het Mirai virus infecteert met name op het internet aangesloten apparaten zoals een DVR, beveiligingscamera of printer.

2. Verander het wachtwoord van de op het internet aangesloten apparaten.
Kies een wachtwoord dat moeilijk te raden is. Als u het huidige wachtwoord niet weet, raadpleeg dan de handleiding.
Door het uitvoeren van deze stappen heeft u toekomstige infecties voorkomen.

3. Herstart de op het internet aangesloten apparaten door deze uit en opnieuw aan te zetten.
Hierna is het Mirai virus verwijderd uit het geheugen van de apparaten.



Nu uw op het internet aangesloten apparaten veilig zijn, zijn de laatste stappen om uw router/modem te beschermen.

4. Reset uw modem/router naar de fabrieksinstellingen. Op kpn.com/reset-kpn-experiabox is beschreven hoe u dit kunt doen voor een Experia Box.

Ga naar hoe reset ik de KPN Experia Box

5. Stel het wachtwoord van uw modem/router in. Op https://www.kpn.com/faq/16176 is beschreven hoe u dit kunt doen voor een Experia Box.
Ga naar WiFi naam en beveiliging wijzigen

Let op: Als toegang op afstand voor een apparaat absoluut noodzakelijk is, stel dan handmatig port forwards in op uw router voor het betreffende apparaat. Op https://forum.kpn.com/internet-9/port-forwarding-upnp-wat-waarom-en-hoe-322560 is beschreven hoe u dit kunt doen voor een Experia Box.
Ga naar Port-forwarding

Wat gebeurt er als ik niets doe?
Het beveiligingsprobleem op uw internetaansluiting vormt een gevaar. Daarom hebben wij uw internetaansluiting in een veilige omgeving (quarantaine) geplaatst. U kunt tijdelijk beperkt gebruikmaken van uw internetaansluiting. Daarom is het van belang om bovenstaande stappen vandaag nog uit te voeren en te reageren door een e-mail terug te sturen naar abuse@kpn.com.

De afdeling Abuse
De afdeling Abuse van KPN handelt veiligheidsincidenten af voor KPN.

Meer informatie

Hebt u nog vragen?
U kunt uw vragen stellen via e-mail op abuse@kpn.com.

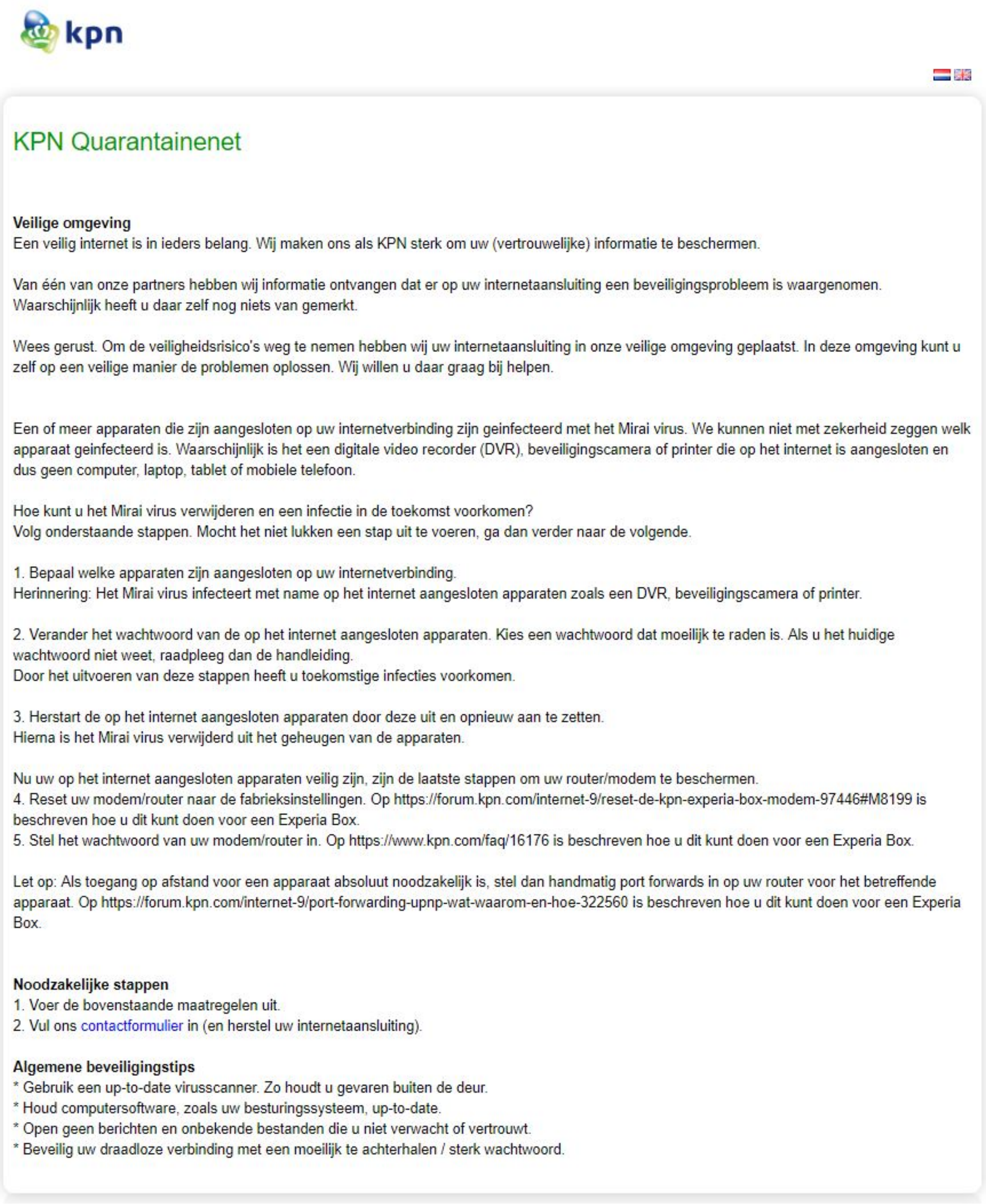
Met vriendelijke groet,

KPN Abuse Team

Wat vindt u van deze e-mail?

KPN B.V. - Postbus 30000 - 2500 GA Den Haag - KvK nr. 27124701

Figure 41 Mirai e-mail notification KPN



KPN Quarantainenet

Veilige omgeving
Een veilig internet is in ieders belang. Wij maken ons als KPN sterk om uw (vertrouwelijke) informatie te beschermen.

Van één van onze partners hebben wij informatie ontvangen dat er op uw internetaansluiting een beveiligingsprobleem is waargenomen. Waarschijnlijk heeft u daar zelf nog niets van gemerkt.

Wees gerust. Om de veiligheidsrisico's weg te nemen hebben wij uw internetaansluiting in onze veilige omgeving geplaatst. In deze omgeving kunt u zelf op een veilige manier de problemen oplossen. Wij willen u daar graag bij helpen.

Een of meer apparaten die zijn aangesloten op uw internetverbinding zijn geinfecteerd met het Mirai virus. We kunnen niet met zekerheid zeggen welk apparaat geinfecteerd is. Waarschijnlijk is het een digitale video recorder (DVR), beveiligingscamera of printer die op het internet is aangesloten en dus geen computer, laptop, tablet of mobiele telefoon.

Hoe kunt u het Mirai virus verwijderen en een infectie in de toekomst voorkomen?
Volg onderstaande stappen. Mocht het niet lukken een stap uit te voeren, ga dan verder naar de volgende.

1. Bepaal welke apparaten zijn aangesloten op uw internetverbinding.
Herinnering: Het Mirai virus infecteert met name op het internet aangesloten apparaten zoals een DVR, beveiligingscamera of printer.

2. Verander het wachtwoord van de op het internet aangesloten apparaten. Kies een wachtwoord dat moeilijk te raden is. Als u het huidige wachtwoord niet weet, raadpleeg dan de handleiding.
Door het uitvoeren van deze stappen heeft u toekomstige infecties voorkomen.

3. Herstart de op het internet aangesloten apparaten door deze uit en opnieuw aan te zetten.
Hierna is het Mirai virus verwijderd uit het geheugen van de apparaten.

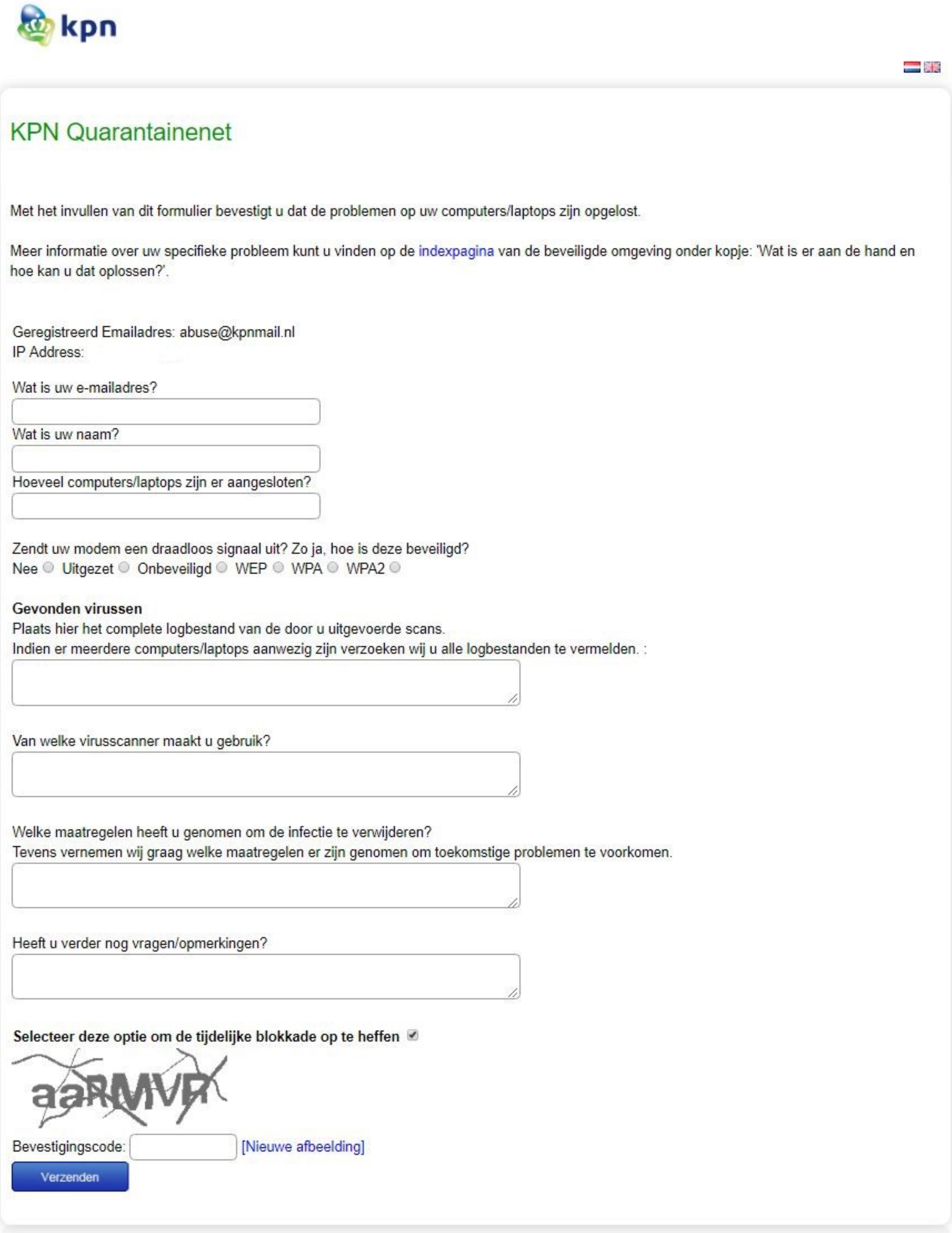
Nu uw op het internet aangesloten apparaten veilig zijn, zijn de laatste stappen om uw router/modem te beschermen.
4. Reset uw modem/router naar de fabrieksinstellingen. Op https://forum.kpn.com/internet-9/reset-de-kpn-experia-box-modem-97446#M8199 is beschreven hoe u dit kunt doen voor een Experia Box.
5. Stel het wachtwoord van uw modem/router in. Op https://www.kpn.com/faq/16176 is beschreven hoe u dit kunt doen voor een Experia Box.

Let op: Als toegang op afstand voor een apparaat absoluut noodzakelijk is, stel dan handmatig port forwards in op uw router voor het betreffende apparaat. Op https://forum.kpn.com/internet-9/port-forwarding-upnp-wat-waarom-en-hoe-322560 is beschreven hoe u dit kunt doen voor een Experia Box.

Noodzakelijke stappen
1. Voer de bovenstaande maatregelen uit.
2. Vul ons contactformulier in (en herstel uw internetaansluiting).

Algemene beveiligingstips
* Gebruik een up-to-date virusscanner. Zo houdt u gevaren buiten de deur.
* Houd computersoftware, zoals uw besturingssysteem, up-to-date.
* Open geen berichten en onbekende bestanden die u niet verwacht of vertrouwt.
* Beveilig uw draadloze verbinding met een moeilijk te achterhalen / sterk wachtwoord.

Figure 42 Mirai landing page KPN - NL



KPN Quarantainenet

Met het invullen van dit formulier bevestigt u dat de problemen op uw computers/laptops zijn opgelost.

Meer informatie over uw specifieke probleem kunt u vinden op de indexpagina van de beveiligde omgeving onder kopje: 'Wat is er aan de hand en hoe kan u dat oplossen?'.

Geregistreerd Emailadres: abuse@kpnmail.nl
IP Address:

Wat is uw e-mailadres?

Wat is uw naam?

Hoeveel computers/laptops zijn er aangesloten?

Zendt uw modem een draadloos signaal uit? Zo ja, hoe is deze beveiligd?
Nee ○ Uitgezet ○ Onbeveiligd ○ WEP ○ WPA ○ WPA2 ○

Gevonden virussen
Plaats hier het complete logbestand van de door u uitgevoerde scans.
Indien er meerdere computers/laptops aanwezig zijn verzoeken wij u alle logbestanden te vermelden. :

Van welke virusscanner maakt u gebruik?

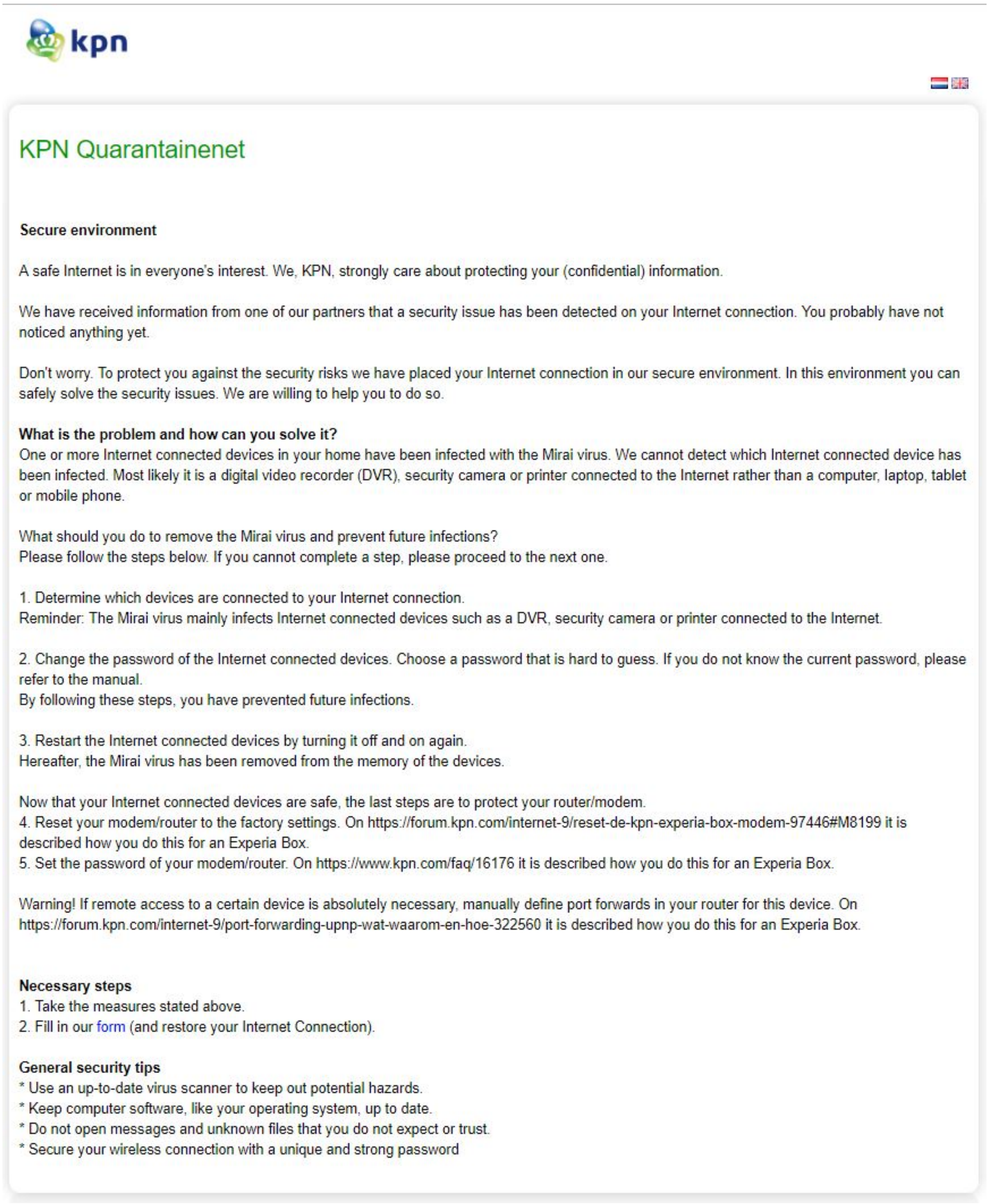
Welke maatregelen heeft u genomen om de infectie te verwijderen?
Tevens vernemen wij graag welke maatregelen er zijn genomen om toekomstige problemen te voorkomen.

Heeft u verder nog vragen/opmerkingen?

Selecteer deze optie om de tijdelijke blokkade op te heffen ☑

Bevestigingscode:

Figure 43 Mirai static contact form KPN - NL



KPN Quarantainenet

Secure environment

A safe Internet is in everyone's interest. We, KPN, strongly care about protecting your (confidential) information.

We have received information from one of our partners that a security issue has been detected on your Internet connection. You probably have not noticed anything yet.

Don't worry. To protect you against the security risks we have placed your Internet connection in our secure environment. In this environment you can safely solve the security issues. We are willing to help you to do so.

What is the problem and how can you solve it?
One or more Internet connected devices in your home have been infected with the Mirai virus. We cannot detect which Internet connected device has been infected. Most likely it is a digital video recorder (DVR), security camera or printer connected to the Internet rather than a computer, laptop, tablet or mobile phone.

What should you do to remove the Mirai virus and prevent future infections?
Please follow the steps below. If you cannot complete a step, please proceed to the next one.

1. Determine which devices are connected to your Internet connection.
Reminder: The Mirai virus mainly infects Internet connected devices such as a DVR, security camera or printer connected to the Internet.

2. Change the password of the Internet connected devices. Choose a password that is hard to guess. If you do not know the current password, please refer to the manual.
By following these steps, you have prevented future infections.

3. Restart the Internet connected devices by turning it off and on again.
Hereafter, the Mirai virus has been removed from the memory of the devices.

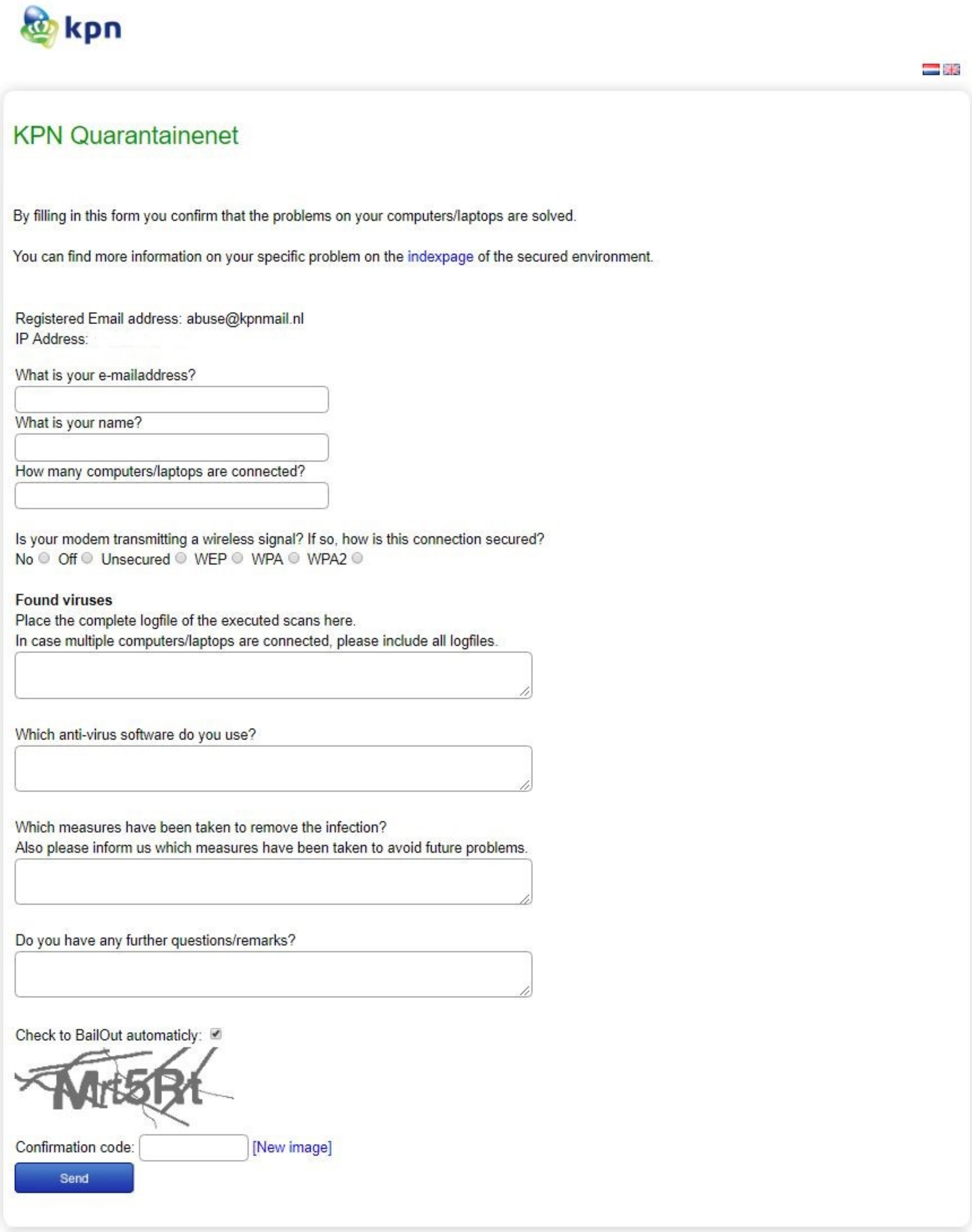
Now that your Internet connected devices are safe, the last steps are to protect your router/modem.
4. Reset your modem/router to the factory settings. On https://forum.kpn.com/internet-9/reset-de-kpn-experia-box-modem-97446#M8199 it is described how you do this for an Experia Box.
5. Set the password of your modem/router. On https://www.kpn.com/faq/16176 it is described how you do this for an Experia Box.

Warning! If remote access to a certain device is absolutely necessary, manually define port forwards in your router for this device. On https://forum.kpn.com/internet-9/port-forwarding-upnp-wat-waarom-en-hoe-322560 it is described how you do this for an Experia Box.

Necessary steps
1. Take the measures stated above.
2. Fill in our form (and restore your Internet Connection).

General security tips
* Use an up-to-date virus scanner to keep out potential hazards.
* Keep computer software, like your operating system, up to date.
* Do not open messages and unknown files that you do not expect or trust.
* Secure your wireless connection with a unique and strong password

Figure 44 Mirai landing page KPN - EN



KPN Quarantainenet

By filling in this form you confirm that the problems on your computers/laptops are solved.

You can find more information on your specific problem on the indexpage of the secured environment.

Registered Email address: abuse@kpnmail.nl
IP Address:

What is your e-mailaddress?

What is your name?

How many computers/laptops are connected?

Is your modem transmitting a wireless signal? If so, how is this connection secured?
No ○ Off ○ Unsecured ○ WEP ○ WPA ○ WPA2 ○

Found viruses
Place the complete logfile of the executed scans here.
In case multiple computers/laptops are connected, please include all logfiles.

Which anti-virus software do you use?

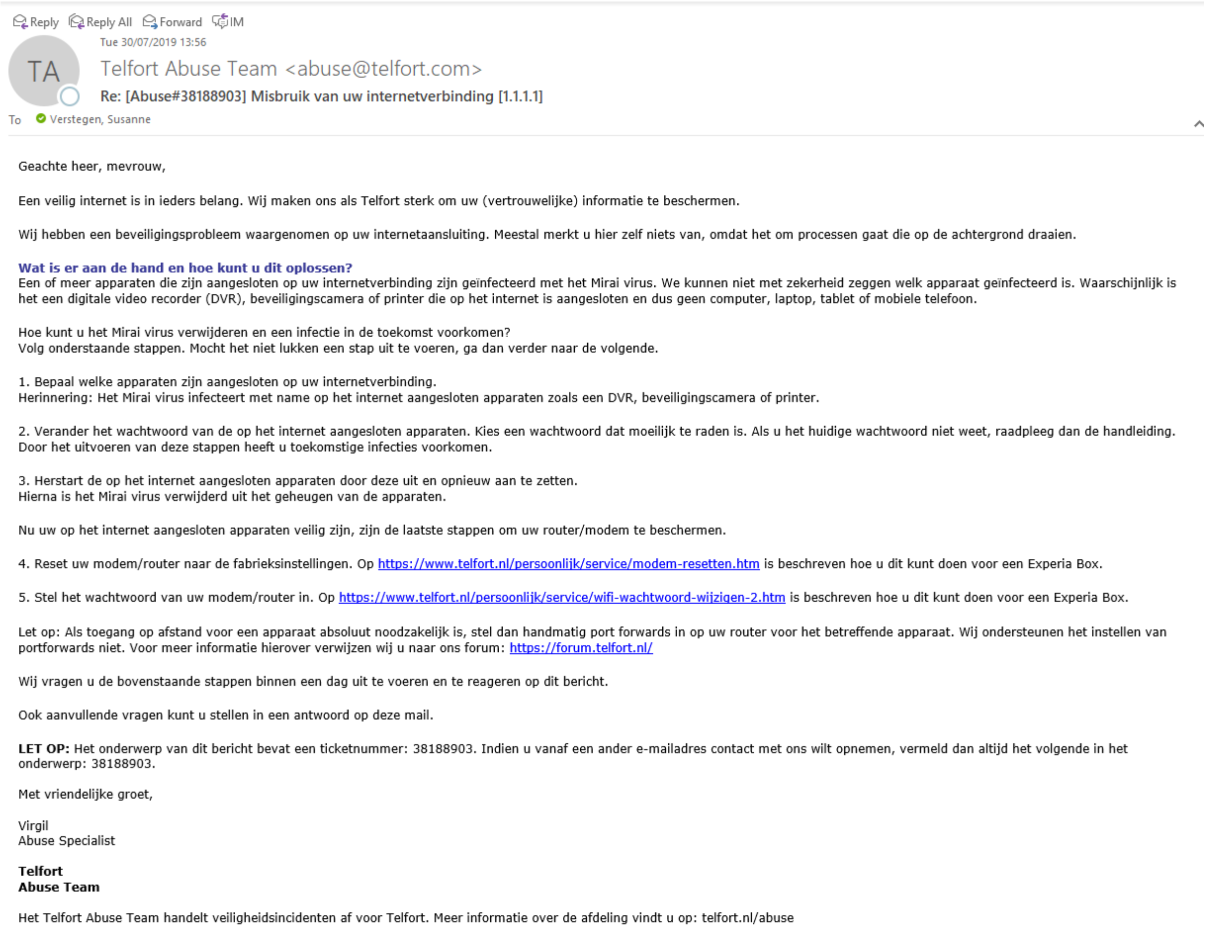
Which measures have been taken to remove the infection?
Also please inform us which measures have been taken to avoid future problems.

Do you have any further questions/remarks?

Check to BailOut automaticly: ☑

Confirmation code: [New image]

Figure 45 Mirai static contact form - EN



Tue 30/07/2019 13:56

Telfort Abuse Team <abuse@telfort.com>

Re: [Abuse#38188903] Misbruik van uw internetverbinding [1.1.1.1]

To Verstegen, Susanne

Geachte heer, mevrouw,

Een veilig internet is in ieders belang. Wij maken ons als Telfort sterk om uw (vertrouwelijke) informatie te beschermen.

Wij hebben een beveiligingsprobleem waargenomen op uw internetaansluiting. Meestal merkt u hier zelf niets van, omdat het om processen gaat die op de achtergrond draaien.

Wat is er aan de hand en hoe kunt u dit oplossen?
Een of meer apparaten die zijn aangesloten op uw internetverbinding zijn geïnfecteerd met het Mirai virus. We kunnen niet met zekerheid zeggen welk apparaat geïnfecteerd is. Waarschijnlijk is het een digitale video recorder (DVR), beveiligingscamera of printer die op het internet is aangesloten en dus geen computer, laptop, tablet of mobiele telefoon.

Hoe kunt u het Mirai virus verwijderen en een infectie in de toekomst voorkomen?
Volg onderstaande stappen. Mocht het niet lukken een stap uit te voeren, ga dan verder naar de volgende.

1. Bepaal welke apparaten zijn aangesloten op uw internetverbinding.
Herinnering: Het Mirai virus infecteert met name op het internet aangesloten apparaten zoals een DVR, beveiligingscamera of printer.

2. Verander het wachtwoord van de op het internet aangesloten apparaten. Kies een wachtwoord dat moeilijk te raden is. Als u het huidige wachtwoord niet weet, raadpleeg dan de handleiding.
Door het uitvoeren van deze stappen heeft u toekomstige infecties voorkomen.

3. Herstart de op het internet aangesloten apparaten door deze uit en opnieuw aan te zetten.
Hierna is het Mirai virus verwijderd uit het geheugen van de apparaten.

Nu uw op het internet aangesloten apparaten veilig zijn, zijn de laatste stappen om uw router/modem te beschermen.

4. Reset uw modem/router naar de fabrieksinstellingen. Op https://www.telfort.nl/persoonlijk/service/modem-resetten.htm is beschreven hoe u dit kunt doen voor een Experia Box.

5. Stel het wachtwoord van uw modem/router in. Op https://www.telfort.nl/persoonlijk/service/wifi-wachtwoord-wijzigen-2.htm is beschreven hoe u dit kunt doen voor een Experia Box.

Let op: Als toegang op afstand voor een apparaat absoluut noodzakelijk is, stel dan handmatig port forwards in op uw router voor het betreffende apparaat. Wij ondersteunen het instellen van portforwards niet. Voor meer informatie hierover verwijzen wij u naar ons forum: https://forum.telfort.nl/

Wij vragen u de bovenstaande stappen binnen een dag uit te voeren en te reageren op dit bericht.

Ook aanvullende vragen kunt u stellen in een antwoord op deze mail.

LET OP: Het onderwerp van dit bericht bevat een ticketnummer: 38188903. Indien u vanaf een ander e-mailadres contact met ons wilt opnemen, vermeld dan altijd het volgende in het onderwerp: 38188903.

Met vriendelijke groet,

Virgil
Abuse Specialist

Telfort
Abuse Team

Het Telfort Abuse Team handelt veiligheidsincidenten af voor Telfort. Meer informatie over de afdeling vindt u op: telfort.nl/abuse

Figure 46 Mirai e-mail notification Telfort - NL

E Background information Shadowserver

CONFIDENTIAL

F Randomization protocol

This appendix describes the randomization protocol that is deployed to allocate detected IPs to a treatment or control group. The following conditions are considered while setting up the protocol:

- The exact sample size (N) is not known in advance;
- We strive for an equal number of IPs in each group;
- Telfort and KPN customers will be treated as two populations. Therefore, both markets have their own protocol.

To assign detected IPs to a group, a list is made for both markets (Telfort and KPN) which determines the sequence of assignment. We choose for a complete random assignment which is a procedure in which each treatment condition contains an equal number of units. The difficulty in creating the assignment lists lies in estimating beforehand how many IPs will have to be assigned: when a list is larger than the number of actual IPs detected, there is still the chance of unequal distribution over the treatment conditions. For that reason, the lists are dynamic: a new list will be added when the previous list is completely used. The new lists will be created with another seed. The complete random assignment is done in R using the package `randomizr`. The following two sections show the used code.

KPN assignment list

```
> install.packages("randomizr")

> set.seed(24)

> Z <- complete_ra(99, num_arms = 3, conditions = c("control","e-m
ail only","loose wg"), check_inputs = TRUE)

#initial list of 99 assignments
#equally distributed over three treatment conditions

> write.table(Z, file="KPN_list.csv",sep=",",row.names=F)


> set.seed(25)

> Z <- complete_ra(33, num_arms = 3, conditions = c("control","e-m
ail only","loose wg"), check_inputs = TRUE)

#second list of 33 assignments
#equally distributed over three treatment conditions

> write.table(Z, file="KPN_list.csv",sep=",",row.names=F)
```

Telfort assignment list

```
> set.seed(48)

> Z <- complete_ra(60, num_arms = 3, conditions = c("contro
l","e-mail only","loose wg"), check_inputs = TRUE)

#initial list of 60 assignments
#equally distributed over three treatment conditions

> write.table(Z, file="Telfort_list.csv",sep=",",row.names=
F)
```

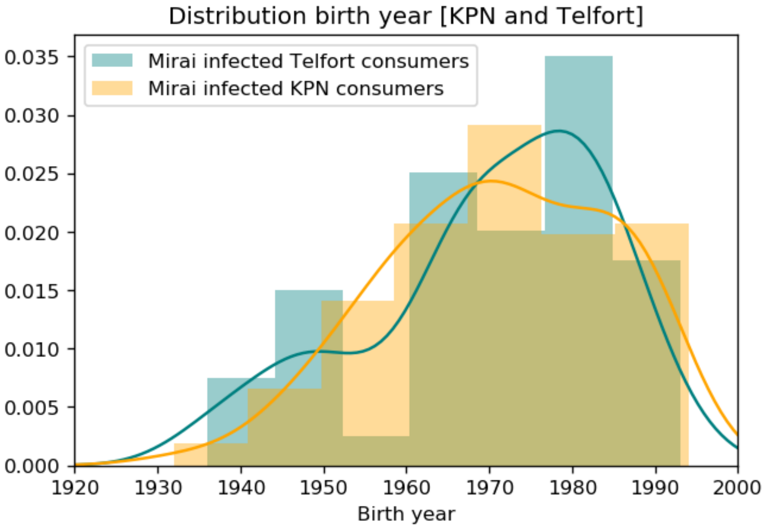
G Interview protocol

#	question	category	stage	Remediation efforts	Reasons of non-compliance	Recurrent notifications	Device type	Customer reaction	E-mail only		Loose walled garden		Control	
									yes	no	yes	no	yes	no
1.1	Do you have time?	introductory						x	2.1	1.2	9.1	1.2	12.1	1.2
1.2	What moment would suit you better?	Closing						x						
2.1	Do you recall receiving the notification?	Transition	Delivery		x				3.1	2.2	-	-	-	
2.2	Is [e-mail address] your correct e-mail address?	Closing			x				2.4	2.3	-	-	-	-
2.3	What is your correct e-mail address?	Closing			x			x			-	-	-	-
2.4	Do you know the possible reason(s) for not receiving or noticing the e-mail?	Transition			x				8.1	-	-	-	-	-
3.1	Have you had the chance to read the notification?	Transition	Attention		x			x	4.1	3.2	-	-	-	-
3.2	What contributed to not reading the e-mail?	key		x	x				8.1		-	-	-	-
4.1	At that moment, did you understand the content of the e-mail?	Transition	Comprehension		x				5.1	4.2	-	-	-	-

4.2	Do you remember what was not clear to you?	Key		x	x				5.1		5.1		-	-
5.1	Have you tried to perform the recommended actions?	Key	Intention	x	x			x	7.1	6.1	7.1	7.1	-	-
6.1	What demotivated you or hold you back?	Key	**Motivation** /Beliefs	x	x				8.1		8.1		-	-
7.1	Did you succeed in performing the recommended actions?	Transition		x	x				7.3	7.2	7.3	7.2	-	-
7.2	What have you tried?	Key	Behavior	x	x		x		7.4		7.4		-	-
7.3	How did you do that?	Key		x	x		x		7.4		7.4		-	-
7.4	Which device(s) have you identified as possibly infected?	Key					x		8.1		8.1		-	-
8.1	What do you think of KPN's service to reach out to infected customers?	Key						x	8.2		8.2		-	-
8.2	How can this service be improved?	Key						x	13.1		13.1		-	-
9.1	Do you recall being placed into quarantine?	transition	Delivery		x				-	-	10.1	9.2	-	-
9.2	Is there a chance another user of your Internet connection has solved the problem?	transition			x				-	-	9.3	-	-	-
9.3	Could I speak to this person?	transition						x	-	-	1.1	-	-	-
10.1	At that moment, did you understand the content of the message that was placed into you browser and/or e-mailed to you?	Key	Comprehension		X				-	-	5.1	4.2	-	-
12.1	Do you recall having installed a new device or switched on a device?	Key							-	-	-	-	12.2	12.4
12.2	Which kind of device was that?	Key					x		-	-	-	-	12.3	

12.3	What have you done with that device after use?	Key	Behavior	x					-	-	-	-	13.1	
12.4	Do you have any devices that are connected to the Internet?	Key					x		-	-	-	-	12.5	12.6
12.5	What devices?	Key					x		-	-	-	-	12.6	
12.6	Could you think of another reason that one of these devices are infected?	Key	Behavior	x					-	-	-	-	13.1	
13.1	Is there anything you like to add or ask?	Closing						x	-	-	-	-	-	

H Age distribution KPN and Telfort infected consumers



I Data exploration

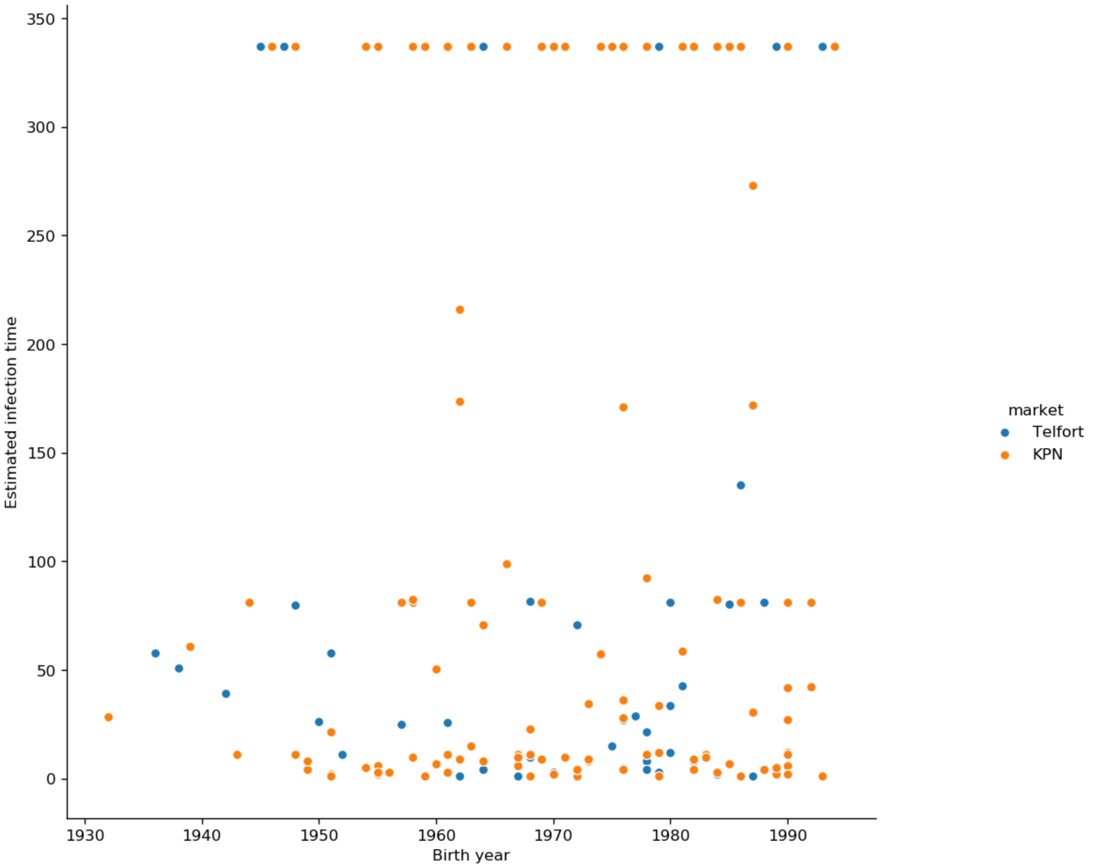


Figure 47 Pairplot all observations, division in market

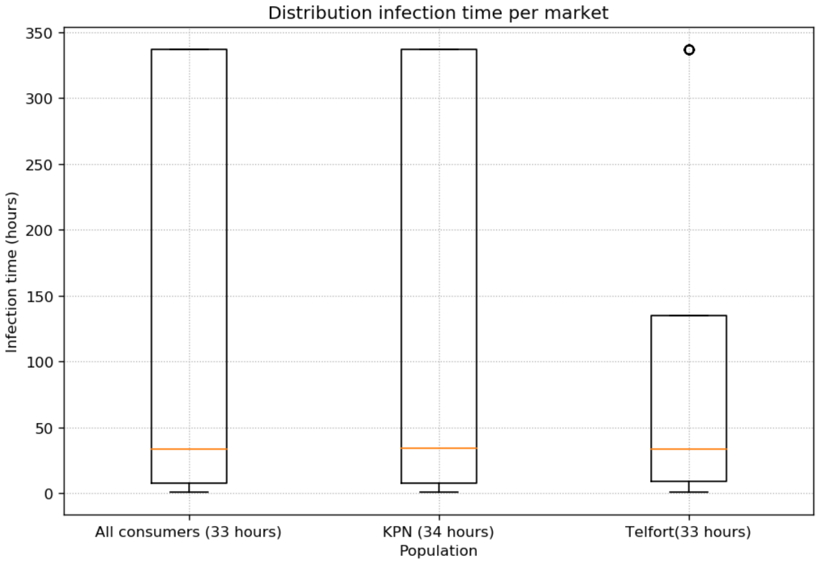


Figure 48 Boxplot all observations per market (median infection time)

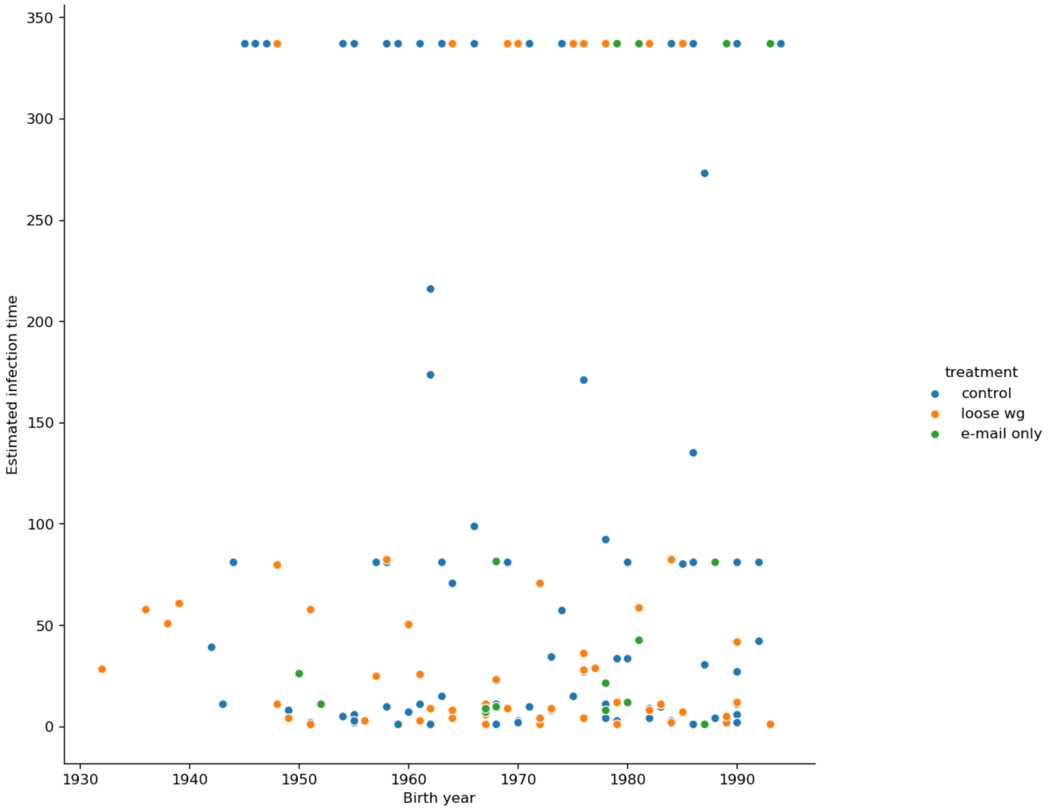


Figure 49 Pairplot all observations, division in treatment

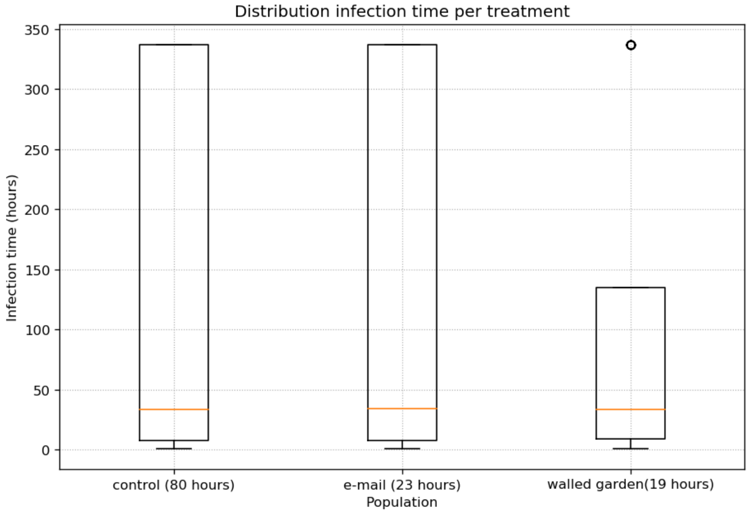


Figure 50 Boxplot all observations per treatment (median infection time)

J Modeling step 1 [all observations]

J.1 Introduction

The complete dataset contains 177 consumers, of which 48 censored (remediation happened after the experiment period). Both female and male are included as variables since there are nine cases in which the gender of the subscriber is unknown. The dummy variables are coded the following:

	Variable ‘female’	Variable ‘male’
Female subscriber	1	0
Male subscriber	0	1
Unknown gender	0	0

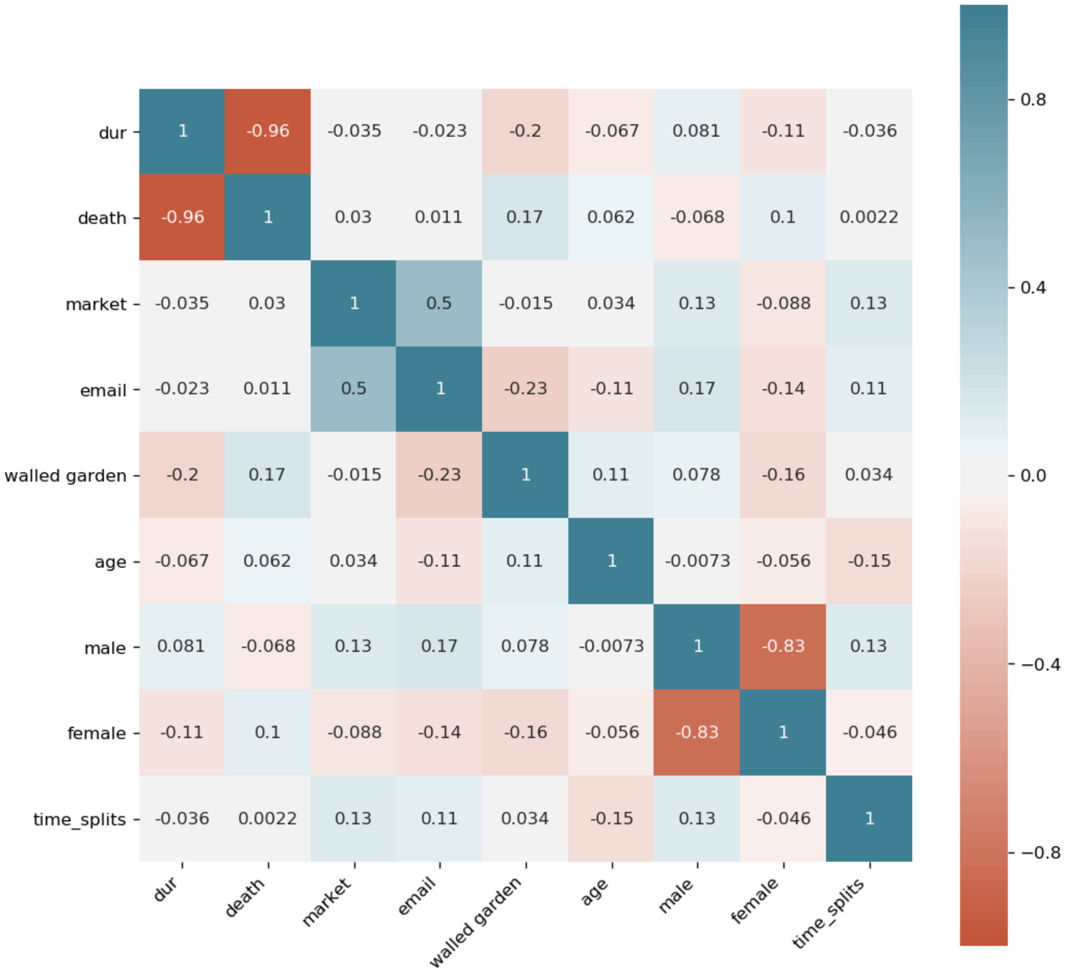
	Variable ‘walled garden’	Variable ‘e-mail
Walled garden notification	1	0
E-mail notification	0	1
No notification (control group)	0	0

	Variable ‘market
KPN consumer	0
Telfort consumer	1

	Variable ‘time_splits’
First detected before June 9th	0
First detected after June 9th	1

J.2 Overview data

Figure 53 shows the correlation of all variables. ‘infection time’ and ‘censored’ are the dependent variables. The other seven variables are independent (possible covariates). The variables market and e-mail are closely linked since the e-mail treatment group only exist of Telfort consumers due to malfunctioning KPN mail server. The correlation between male and female is high (only eleven cases of unknown gender). To avoid multicollinearity, only one of these two variables is chosen. The variables are separately modeled to check which more reliable in step 0 (cannot be modeled as one dummy variable due to the unknown gender cases).



	dur	death	market	email	walled garden	age	male	female	time_splits
dur	1	-0.96	-0.035	-0.023	-0.2	-0.067	0.081	-0.11	-0.036
death	-0.96	1	0.03	0.011	0.17	0.062	-0.068	0.1	0.0022
market	-0.035	0.03	1	0.5	-0.015	0.034	0.13	-0.088	0.13
email	-0.023	0.011	0.5	1	-0.23	-0.11	0.17	-0.14	0.11
walled garden	-0.2	0.17	-0.015	-0.23	1	0.11	0.078	-0.16	0.034
age	-0.067	0.062	0.034	-0.11	0.11	1	-0.0073	-0.056	-0.15
male	0.081	-0.068	0.13	0.17	0.078	-0.0073	1	-0.83	0.13
female	-0.11	0.1	-0.088	-0.14	-0.16	-0.056	-0.83	1	-0.046
time_splits	-0.036	0.0022	0.13	0.11	0.034	-0.15	0.13	-0.046	1

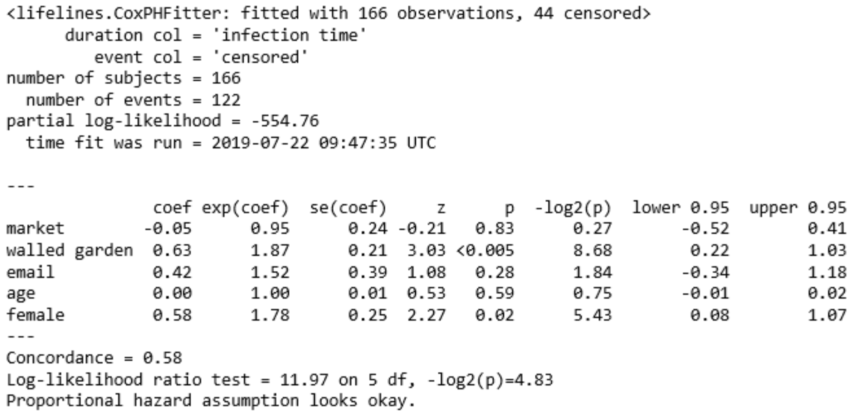
Figure 51 Correlation potential remediation drivers (all observations)

J.3 Cox modeling steps

Step 0

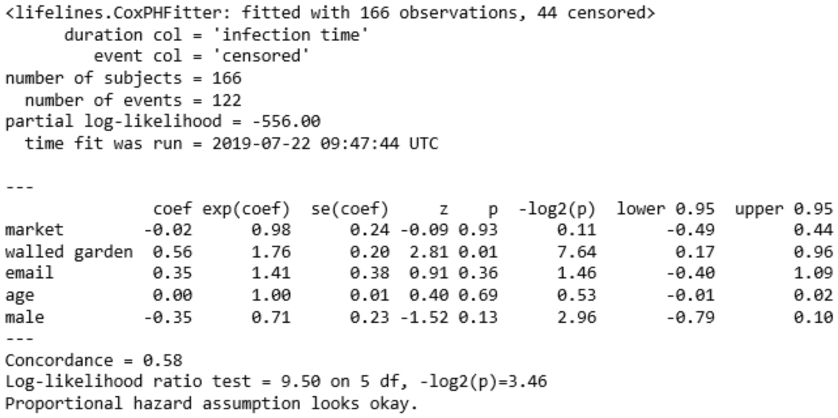
First, we check if ‘female’ or ‘male’ is a better covariate.

Including female:



```
<lifelines.CoxPHFitter: fitted with 166 observations, 44 censored>
      duration col = 'infection time'
         event col = 'censored'
number of subjects = 166
  number of events = 122
partial log-likelihood = -554.76
  time fit was run = 2019-07-22 09:47:35 UTC

---
               coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market        -0.05      0.95      0.24 -0.21   0.83      0.27       -0.52        0.41
walled garden  0.63      1.87      0.21  3.03 <0.005      8.68        0.22        1.03
email          0.42      1.52      0.39  1.08   0.28      1.84       -0.34        1.18
age            0.00      1.00      0.01  0.53   0.59      0.75       -0.01        0.02
female         0.58      1.78      0.25  2.27   0.02      5.43        0.08        1.07
---
Concordance = 0.58
Log-likelihood ratio test = 11.97 on 5 df, -log2(p)=4.83
Proportional hazard assumption looks okay.
```

Including male:



```
<lifelines.CoxPHFitter: fitted with 166 observations, 44 censored>
      duration col = 'infection time'
         event col = 'censored'
number of subjects = 166
  number of events = 122
partial log-likelihood = -556.00
  time fit was run = 2019-07-22 09:47:44 UTC

---
               coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
market        -0.02      0.98      0.24 -0.09 0.93      0.11       -0.49        0.44
walled garden  0.56      1.76      0.20  2.81 0.01      7.64        0.17        0.96
email          0.35      1.41      0.38  0.91 0.36      1.46       -0.40        1.09
age            0.00      1.00      0.01  0.40 0.69      0.53       -0.01        0.02
male          -0.35      0.71      0.23 -1.52 0.13      2.96       -0.79        0.10
---
Concordance = 0.58
Log-likelihood ratio test = 9.50 on 5 df, -log2(p)=3.46
Proportional hazard assumption looks okay.
```

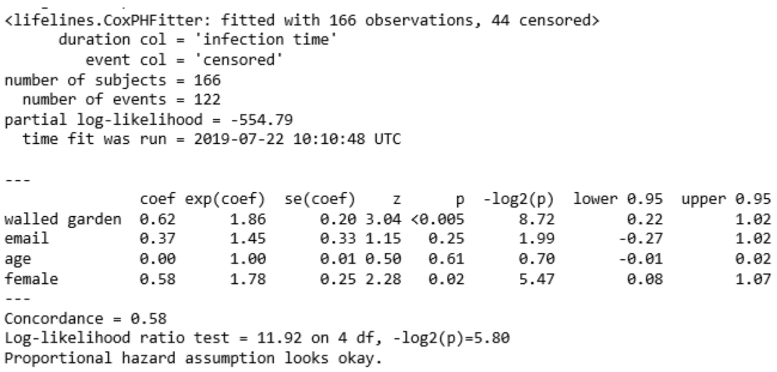
The model including ‘female’ has higher reliable parameters and a better overall fit (the partial log-likelihood is higher for equal degrees of freedom).

In the next steps, the variable ‘male’ is excluded due to the high covariance between this variable and ‘female’. Due to this exclusion, the coding of the dummy variable ‘female’ is changed. 1 = female subscriber, and 0 = male subscriber and subscribers of unknown gender.

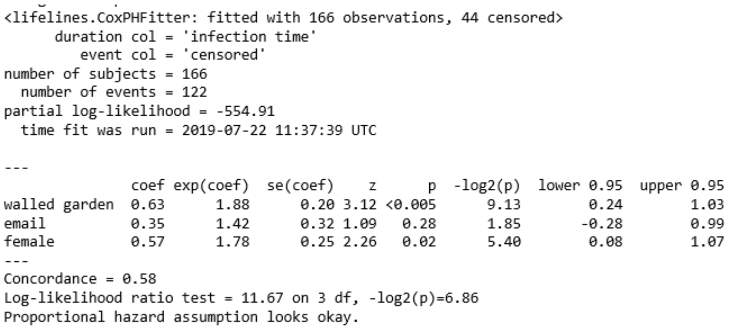
Step 1

The market variable is the least reliable and excluded

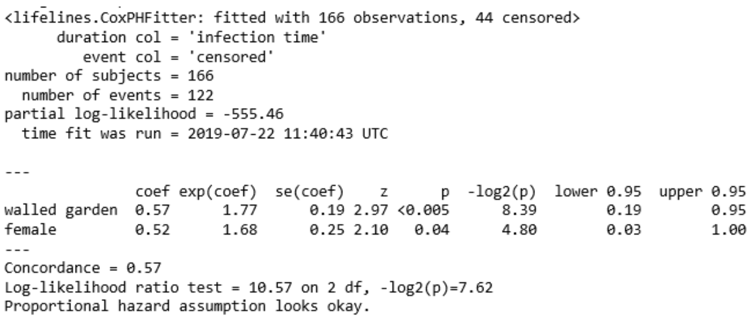
Modeling the Cox model for [age, female, walled garden, e-mail]

<table>
<tr>
<td></td>
<td>

<pre>
<lifelines.CoxPHFitter: fitted with 166 observations, 44 censored>
      duration col = 'infection time'
         event col = 'censored'
number of subjects = 166
  number of events = 122
partial log-likelihood = -554.79
  time fit was run = 2019-07-22 10:10:48 UTC

---
               coef exp(coef) se(coef)   z   p -log2(p) lower 0.95 upper 0.95
walled garden 0.62   1.86   0.20 3.04 <0.005   8.72   0.22   1.02
email   0.37   1.45   0.33 1.15   0.25   1.99   -0.27   1.02
age   0.00   1.00   0.01 0.50   0.61   0.70   -0.01   0.02
female   0.58   1.78   0.25 2.28   0.02   5.47   0.08   1.07

Concordance = 0.58
Log-likelihood ratio test = 11.92 on 4 df, -log2(p)=5.80
Proportional hazard assumption looks okay.
</pre>
<p>The difference in partial log-likelihood (LL) = -0,03</p>
<p>Log-Likelihood Ratio Statistic (LRS) = 0,06</p>
<p>Consulting the chi-square distribution for 0,06 on 1 df: p&gt;0,75</p>
<p>Model 1 is better than model 0</p>
</td>
</tr>
<tr>
<td>Step 2</td>
<td>
<p>The age variable is the least reliable and excluded</p>
<p>Modeling the Cox model for [female, walled garden, e-mail]</p>

<pre>
<lifelines.CoxPHFitter: fitted with 166 observations, 44 censored>
      duration col = 'infection time'
         event col = 'censored'
number of subjects = 166
  number of events = 122
partial log-likelihood = -554.91
  time fit was run = 2019-07-22 11:37:39 UTC

---
               coef exp(coef) se(coef)   z   p -log2(p) lower 0.95 upper 0.95
walled garden 0.63   1.88   0.20 3.12 <0.005   9.13   0.24   1.03
email   0.35   1.42   0.32 1.09   0.28   1.85   -0.28   0.99
female   0.57   1.78   0.25 2.26   0.02   5.40   0.08   1.07

Concordance = 0.58
Log-likelihood ratio test = 11.67 on 3 df, -log2(p)=6.86
Proportional hazard assumption looks okay.
</pre>
<p>The difference in partial log-likelihood (LL) = -0,12</p>
<p>Log-Likelihood Ratio Statistic (LRS) = 0,24</p>
<p>Consulting the chi-square distribution for 0,24 on 1 df: p&gt;0,50</p>
<p>Model 2 is better than model 1</p>
</td>
</tr>
<tr>
<td>Step 3</td>
<td>
<p>The e-mail variable is the least reliable and excluded</p>
<p>Modeling the Cox model for [female, walled garden]</p>

<pre>
<lifelines.CoxPHFitter: fitted with 166 observations, 44 censored>
      duration col = 'infection time'
         event col = 'censored'
number of subjects = 166
  number of events = 122
partial log-likelihood = -555.46
  time fit was run = 2019-07-22 11:40:43 UTC

---
               coef exp(coef) se(coef)   z   p -log2(p) lower 0.95 upper 0.95
walled garden 0.57   1.77   0.19 2.97 <0.005   8.39   0.19   0.95
female   0.52   1.68   0.25 2.10   0.04   4.80   0.03   1.00

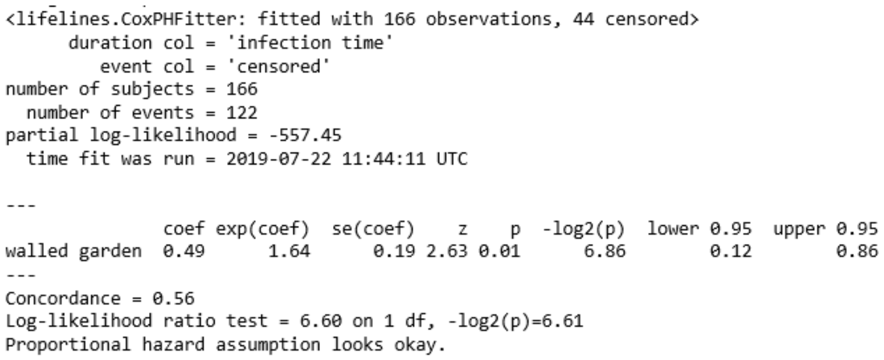
Concordance = 0.57
Log-likelihood ratio test = 10.57 on 2 df, -log2(p)=7.62
Proportional hazard assumption looks okay.
</pre>
<p>The difference in partial log-likelihood (LL) = -0,55</p>
<p>Log-Likelihood Ratio Statistic (LRS) = 1,1</p>
<p>Consulting the chi-square distribution for 1,1 on 1 df: p&gt;0,25</p>
</td>
</tr>
</table>

Model 3 is better than model 2

Step 4

The female variable is the least reliable and excluded

Modeling the Cox model for [walled garden]



```
<lifelines.CoxPHFitter: fitted with 166 observations, 44 censored>
      duration col = 'infection time'
         event col = 'censored'
number of subjects = 166
  number of events = 122
partial log-likelihood = -557.45
  time fit was run = 2019-07-22 11:44:11 UTC

---
               coef exp(coef)  se(coef)    z    p  -log2(p)  lower 0.95  upper 0.95
walled garden  0.49      1.64      0.19 2.63 0.01      6.86        0.12        0.86
---
Concordance = 0.56
Log-likelihood ratio test = 6.60 on 1 df, -log2(p)=6.61
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -1,99

Log-Likelihood Ratio Statistic (LRS) = 3,98

Consulting the chi-square distribution for 3,98 on 1 df: p<0,05

Model 3 is better than model 4

Step 5

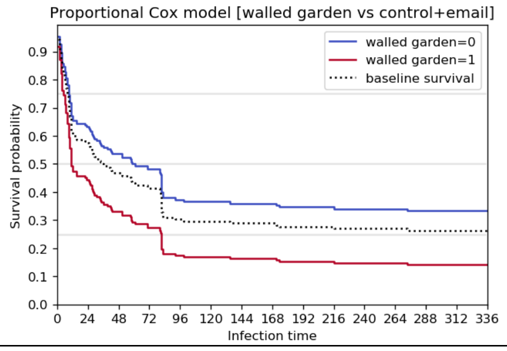
Model 3 is best of all models.

When comparing model 3 with a trivial model, the LRS is 10,57 for 2 degrees of freedom: p<0,01

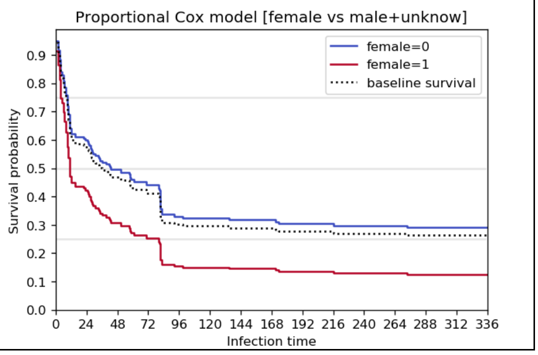
Model 3 is better than a model without covariates.

Model 3 is accepted

Step 6



Proportional Cox model [walled garden vs control+email]



Proportional Cox model [female vs male+unknow]

J.4 AFT modeling steps

Step 0

Prior to estimating a model, we must find the best fitting distribution for the survival curve. Figure 54 shows the distribution fits and the Log-Likelihood (LL) of that function compared to the null distributions. Since the data and the number of parameters are for all these five distributions the same, except for the exponential distribution, we can directly compare the LL estimates of these five. The difference in one degree of freedom does not compromise for the low LL of the exponential distribution. When comparing the rest, we can conclude the LogNormal distribution has the best goodness of fit.

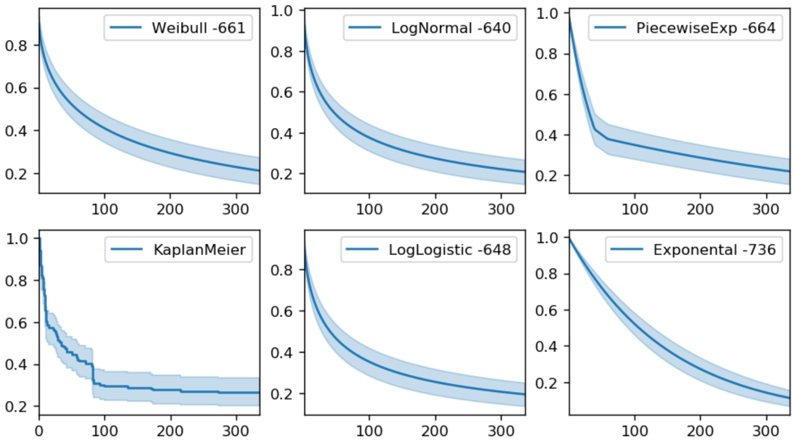


Figure 52 Distribution fits for the survival curve of all observations

Figure 55 shows the quantile-quantile (Q-Q) plot to compare the fitted LogNormal distribution with the empirical distribution. The Q-Q-plot shows that until 81 hours, the data has quite the same shape as the fitted LogNormal distribution. The empirical distribution is a bit more concentrated than the fitted distribution. Then there is a spike of identical values of 81/82 hours (horizontal line of dots). This can also be seen in the Kaplan-Meier plot in section 6.3, which shows a drop around this time. After this spike, the dots form a steep vertical line, which indicates there is a gap in values.

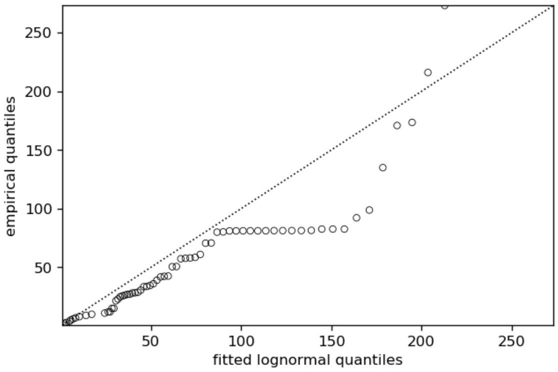
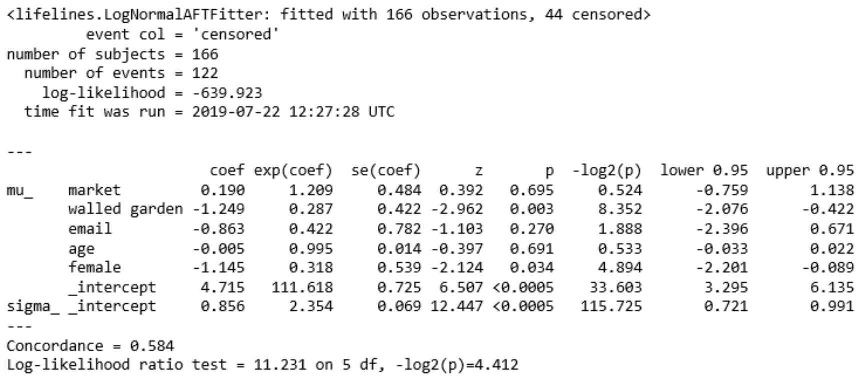


Figure 53 Q-Q plot LogNormal distribution

Step 1

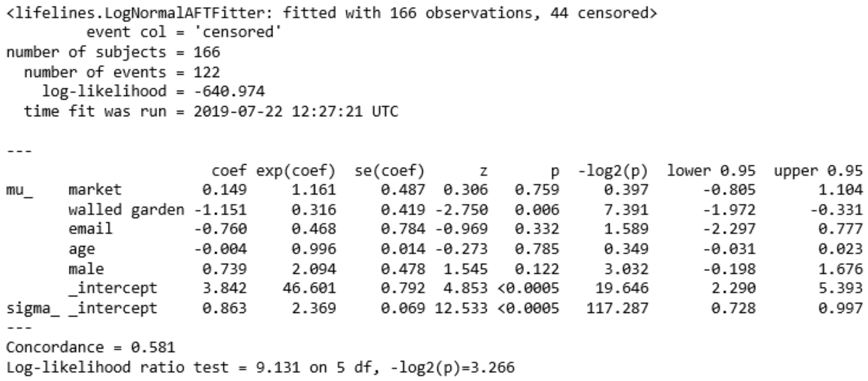
Similar to the Cox modeling in the previous section, we first estimate two models (including female and including male) so we can decide which variable to continue with.

Including female:



```
<lifelines.LogNormalAFTFitter: fitted with 166 observations, 44 censored>
         event col = 'censored'
number of subjects = 166
  number of events = 122
    log-likelihood = -639.923
  time fit was run = 2019-07-22 12:27:28 UTC

---
                       coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market         0.190     1.209     0.484  0.392   0.695     0.524      -0.759       1.138
       walled garden -1.249     0.287     0.422 -2.962   0.003     8.352      -2.076      -0.422
       email         -0.863     0.422     0.782 -1.103   0.270     1.888      -2.396       0.671
       age           -0.005     0.995     0.014 -0.397   0.691     0.533      -0.033       0.022
       female        -1.145     0.318     0.539 -2.124   0.034     4.894      -2.201      -0.089
       _intercept     4.715   111.618     0.725  6.507 <0.0005    33.603       3.295       6.135
sigma_ _intercept     0.856     2.354     0.069 12.447 <0.0005   115.725       0.721       0.991
---
Concordance = 0.584
Log-likelihood ratio test = 11.231 on 5 df, -log2(p)=4.412
```

Including male:



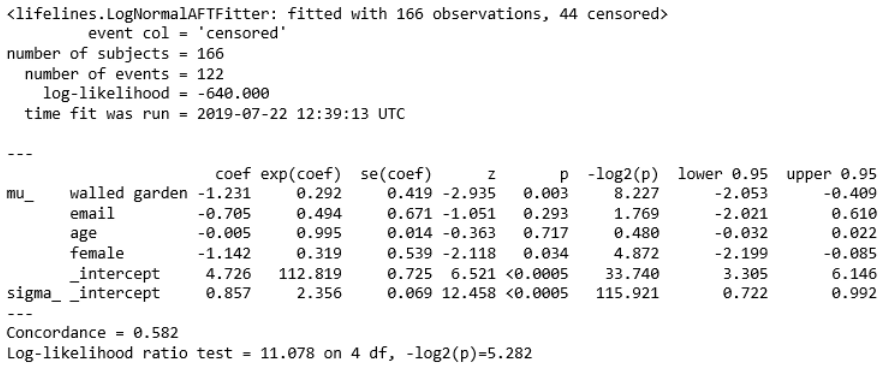
```
<lifelines.LogNormalAFTFitter: fitted with 166 observations, 44 censored>
         event col = 'censored'
number of subjects = 166
  number of events = 122
    log-likelihood = -640.974
  time fit was run = 2019-07-22 12:27:21 UTC

---
                       coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market         0.149     1.161     0.487  0.306   0.759     0.397      -0.805       1.104
       walled garden -1.151     0.316     0.419 -2.750   0.006     7.391      -1.972      -0.331
       email         -0.760     0.468     0.784 -0.969   0.332     1.589      -2.297       0.777
       age           -0.004     0.996     0.014 -0.273   0.785     0.349      -0.031       0.023
       male           0.739     2.094     0.478  1.545   0.122     3.032      -0.198       1.676
       _intercept     3.842    46.601     0.792  4.853 <0.0005    19.646       2.290       5.393
sigma_ _intercept     0.863     2.369     0.069 12.533 <0.0005   117.287       0.728       0.997
---
Concordance = 0.581
Log-likelihood ratio test = 9.131 on 5 df, -log2(p)=3.266
```

The model including 'female' has higher reliable parameters and a better overall fit (the log-likelihood is higher for equal degrees of freedom). Similar to the Cox model, the next steps will include 'female'. (Dummy coding: 1 = female subscriber, and 0 = male subscriber and subscribers of unknown gender)

Step 2

The market variable is the least reliable and excluded

Modeling the AFT LogNormal model for [age, female, walled garden, e-mail]



```
<lifelines.LogNormalAFTFitter: fitted with 166 observations, 44 censored>
         event col = 'censored'
number of subjects = 166
  number of events = 122
    log-likelihood = -640.000
  time fit was run = 2019-07-22 12:39:13 UTC

---
                       coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    walled garden -1.231     0.292     0.419 -2.935   0.003     8.227      -2.053      -0.409
       email         -0.705     0.494     0.671 -1.051   0.293     1.769      -2.021       0.610
       age           -0.005     0.995     0.014 -0.363   0.717     0.480      -0.032       0.022
       female        -1.142     0.319     0.539 -2.118   0.034     4.872      -2.199      -0.085
       _intercept     4.726   112.819     0.725  6.521 <0.0005    33.740       3.305       6.146
sigma_ _intercept     0.857     2.356     0.069 12.458 <0.0005   115.921       0.722       0.992
---
Concordance = 0.582
Log-likelihood ratio test = 11.078 on 4 df, -log2(p)=5.282
```

The difference in partial log-likelihood (LL) = -0,077

Log-Likelihood Ratio Statistic (LRS) = 0,154

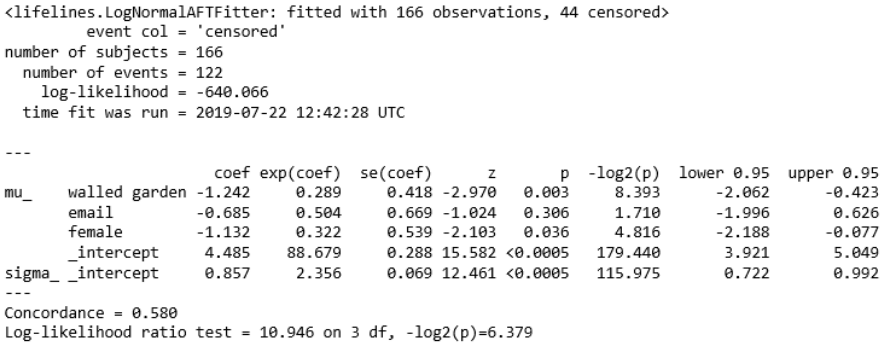
Consulting the chi-square distribution for 0,154 on 1 df: p>0,50

Model 2 is better than model 1

Step 3

The age variable is the least reliable and excluded

Modeling the AFT LogNormal model for [female, walled garden, e-mail]



```
<lifelines.LogNormalAFTFitter: fitted with 166 observations, 44 censored>
         event col = 'censored'
number of subjects = 166
  number of events = 122
    log-likelihood = -640.066
  time fit was run = 2019-07-22 12:42:28 UTC

---
                        coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    walled garden -1.242     0.289     0.418 -2.970  0.003     8.393      -2.062      -0.423
       email         -0.685     0.504     0.669 -1.024  0.306     1.710      -1.996       0.626
       female        -1.132     0.322     0.539 -2.103  0.036     4.816      -2.188      -0.077
       _intercept     4.485    88.679     0.288 15.582 <0.0005   179.440       3.921       5.049
sigma_ _intercept     0.857     2.356     0.069 12.461 <0.0005   115.975       0.722       0.992
---
Concordance = 0.580
Log-likelihood ratio test = 10.946 on 3 df, -log2(p)=6.379
```

The difference in partial log-likelihood (LL) = -0,066

Log-Likelihood Ratio Statistic (LRS) = 0,132

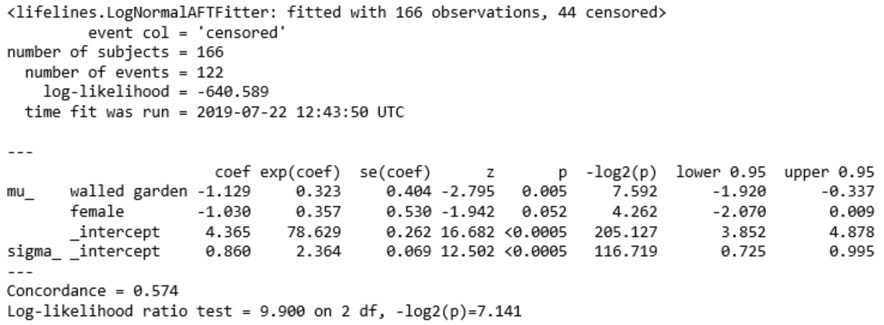
Consulting the chi-square distribution for 0,132 on 1 df: p>0,50

Model 3 is better than model 2

Step 4

The e-mail variable is the least reliable and excluded

Modeling the AFT LogNormal model for [female, walled garden]



```
<lifelines.LogNormalAFTFitter: fitted with 166 observations, 44 censored>
         event col = 'censored'
number of subjects = 166
  number of events = 122
    log-likelihood = -640.589
  time fit was run = 2019-07-22 12:43:50 UTC

---
                        coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    walled garden -1.129     0.323     0.404 -2.795  0.005     7.592      -1.920      -0.337
       female        -1.030     0.357     0.530 -1.942  0.052     4.262      -2.070       0.009
       _intercept     4.365    78.629     0.262 16.682 <0.0005   205.127       3.852       4.878
sigma_ _intercept     0.860     2.364     0.069 12.502 <0.0005   116.719       0.725       0.995
---
Concordance = 0.574
Log-likelihood ratio test = 9.900 on 2 df, -log2(p)=7.141
```

The difference in partial log-likelihood (LL) = -0,523

Log-Likelihood Ratio Statistic (LRS) = 1,046

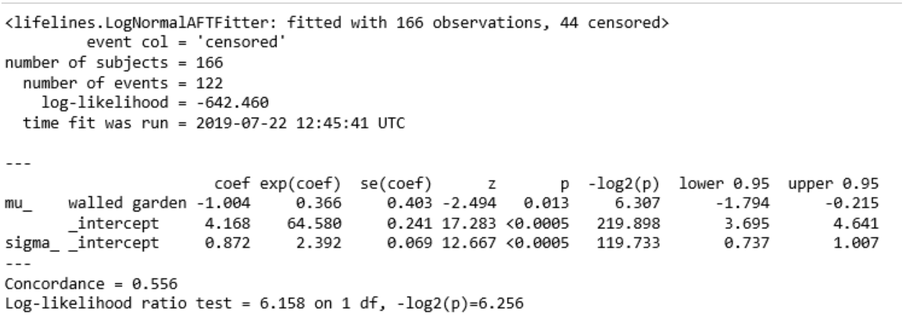
Consulting the chi-square distribution for 1,046 on 1 df: p>0,525

Model 4 is better than model 3

Step 5

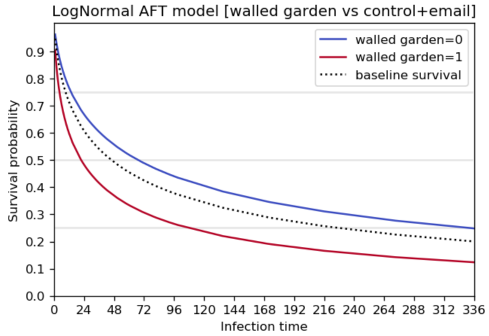
The female variable is the least reliable and excluded

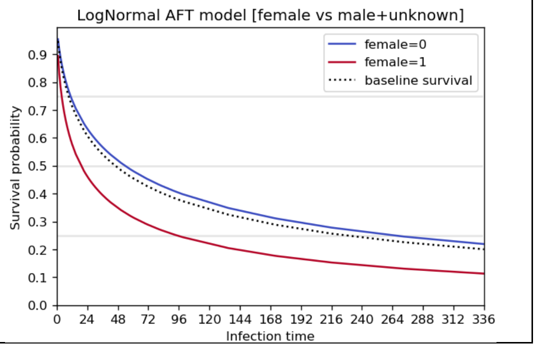
Modeling the AFT LogNormal model for [walled garden]



```
<lifelines.LogNormalAFTFitter: fitted with 166 observations, 44 censored>
         event col = 'censored'
number of subjects = 166
  number of events = 122
    log-likelihood = -642.460
  time fit was run = 2019-07-22 12:45:41 UTC

---
                        coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    walled garden -1.004     0.366     0.403 -2.494  0.013     6.307      -1.794      -0.215
       _intercept     4.168    64.580     0.241 17.283 <0.0005   219.898       3.695       4.641
sigma_ _intercept     0.872     2.392     0.069 12.667 <0.0005   119.733       0.737       1.007
---
Concordance = 0.556
Log-likelihood ratio test = 6.158 on 1 df, -log2(p)=6.256
```

The difference in partial log-likelihood (LL) = -1,871

<table>
<tr><td></td><td>Log-Likelihood Ratio Statistic (LRS) = 3,742
Consulting the chi-square distribution for 3,742 on 1 df: p=0,053
Model 4 is better than model 5 under the significance level of 5,3%</td></tr>
<tr><td>Step 6</td><td>Model 4 is best of all models.
When comparing model 4 with a trivial model, the LRS is 9,9 for 2 degrees of freedom: p<0,05 (p=0,002)
Model 4 is better than a model without covariates.
Model 4 is accepted</td></tr>
<tr><td>Step 7</td><td>
LogNormal AFT model [walled garden vs control+email]


LogNormal AFT model [female vs male+unknown]
</td></tr>
</table>

K Modeling step 2 [interviewed consumers]

K.1 Introduction

Of all the subjects in the experiment, 99 consumers were interviewed. Seven of these consumers had a business account, two consumers were reinfected and one customer had a business account and was reinfected. We exclude these customers from the dataset, resulting in a dataset of 89 entries, of which 22 are censored.

Through the interviews, we obtained more information that may have explanatory value for infection time. This information is translated into several dummy variables. Two dummy variables concern two stages of the theoretical framework: awareness and behavior. Comprehension, intention and compliance are not included since these stages are only of interest in the e-mail and walled garden group. This appendix, therefore, excludes these stages. Appendix L models the process excluding the control group so that the other stages can be included.

The two dummy variables are defined and coded as the following:

Aware of notification (awareness): whether the interviewed consumer has received a notification and is aware of the content.

	Variable 'aware of notification':
Consumers in the control group; consumers in the e-mail and walled group who have not seen or read the notification	0
Consumers in the e-mail and walled group who have seen and read the notification	1

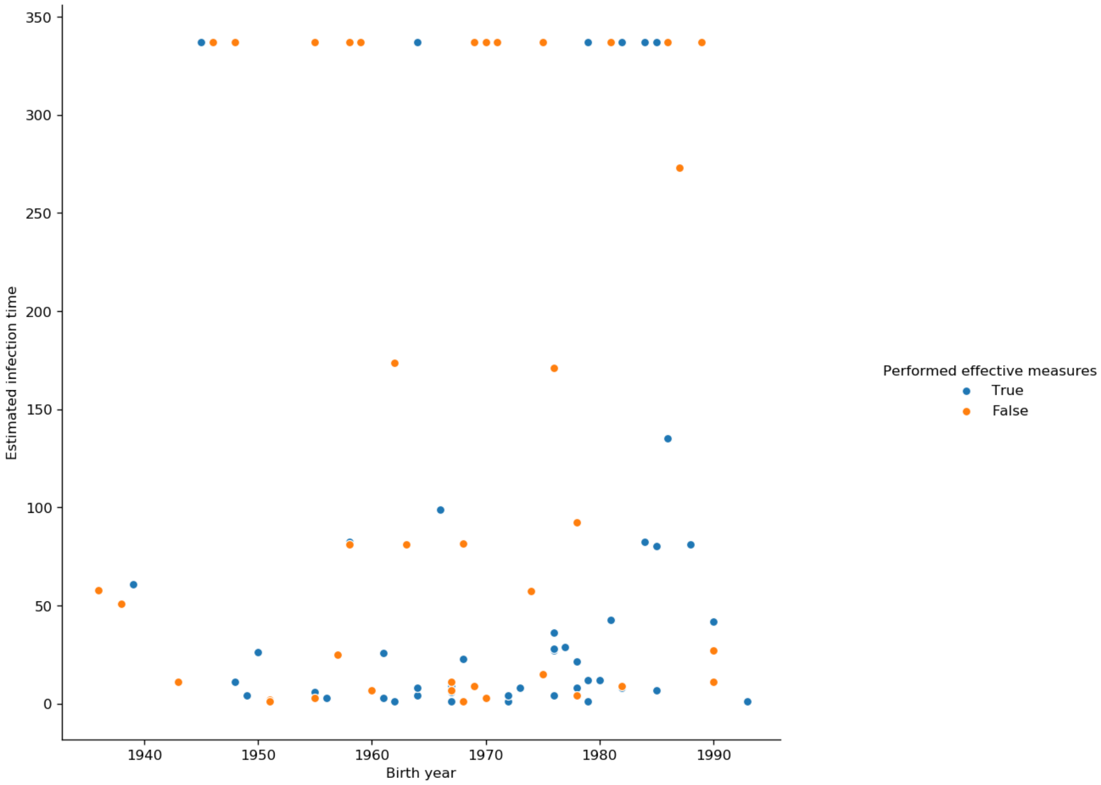
Right measures (behavior): whether the interviewed consumer has performed effective remediation measures. These actions may differ from the recommended actions in the notification. Section 8.1 presents the rules that determine whether remediation actions are considered 'effective'.

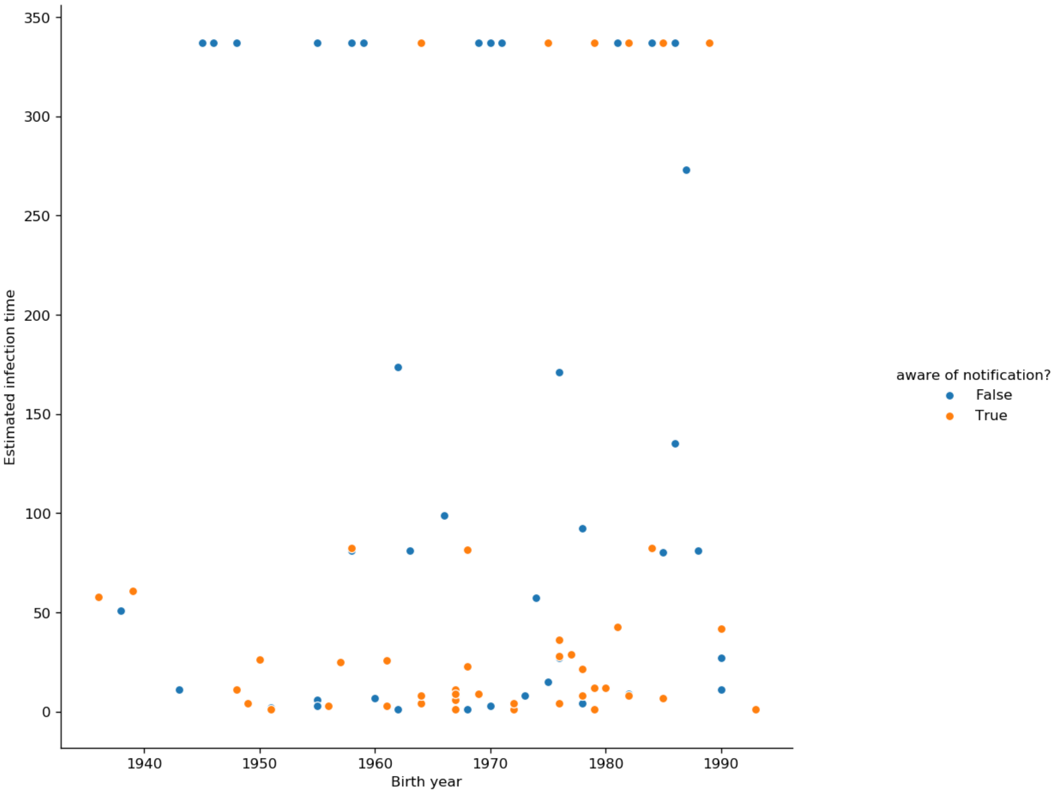
	Variable 'Right measures':
Consumers who haven't performed effective measures to remediate Mirai	0
Consumers who have performed effective measures to remediate Mirai	1

Other data obtained through the interviews is the device types that consumers have identified as infected. The pie charts in chapter 5 visualize the ratios of device types. However, these pie charts do not take into that some consumers have identified several devices as possible infected. Since we don't know the precise device, we collect these devices under the variable 'multiple'. The NAS and Rasberry Pi devices are collected under the variable 'home automation' (modeled as 'home'). This data exists of 39 home automation devices, 23 camera's, 11 instances of multiple possibly infected devices, 11 unknown device types, 3 printers and 2 routers. Due to their low occurrence, printers and routers are not modeled asa variable. All other device types are modeled as dummy variables:

	Variable 'home'	Variable 'camera	Variable 'multiple'	Variable 'unknown'
Consumers who have identified a NAS or Rasberry Pi as the infected device	1	0	0	0
Consumers who have identified an IP camera as the infected device	0	1	0	0
Consumers who have identified multiple IoT devices as possibly infected	0	0	1	0
Consumers who were not able to identify an IoT device in their network	0	0	0	1
Consumers who have identified a printer or router as the infected device (rest group)	0	0	0	0

K.2 Overview data





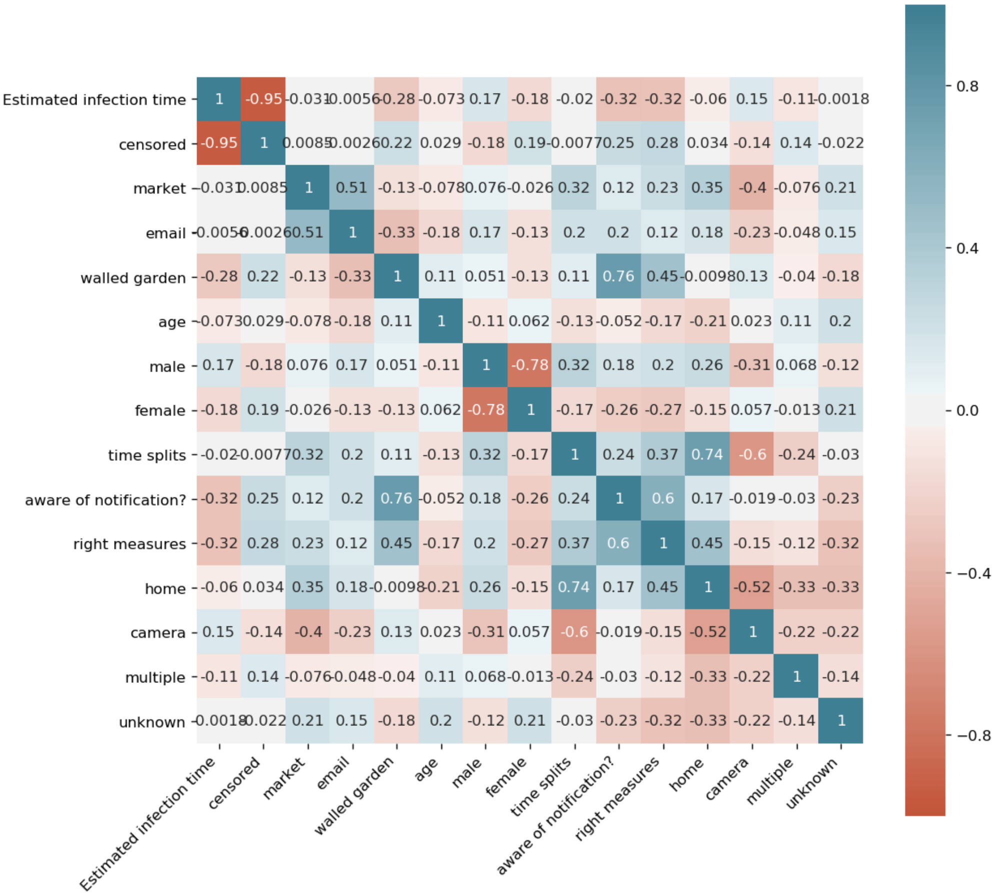


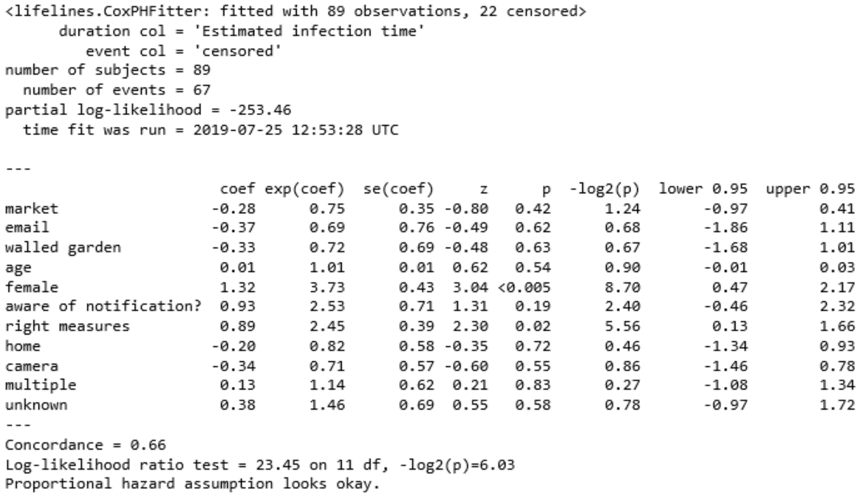
Figure 54 Correlation potential remediation drivers (interviewed consumers)

K.3 Cox modeling steps

Step 0

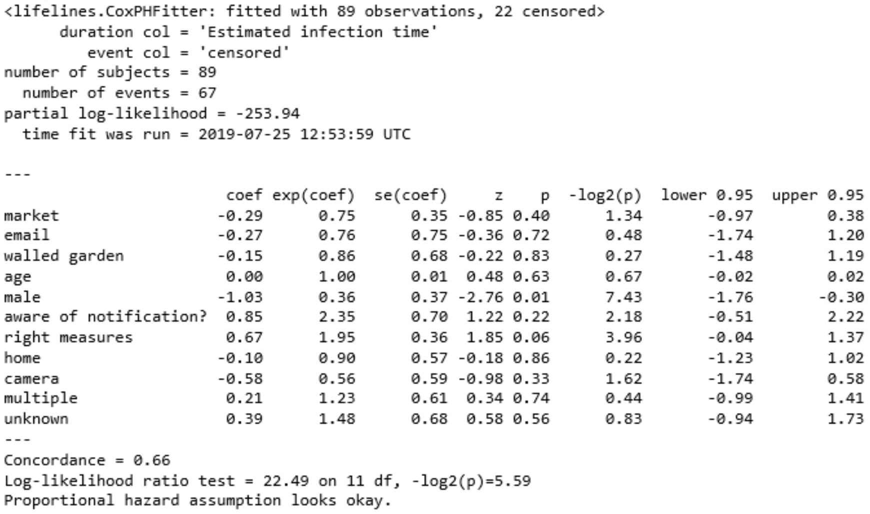
For the same reasons as addressed in appendix J.2, we first check if ‘female’ or ‘male’ is a better covariate to determine which of the two to include.

Including female:



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -253.46
  time fit was run = 2019-07-25 12:53:28 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                 -0.28      0.75      0.35 -0.80   0.42      1.24       -0.97        0.41
email                  -0.37      0.69      0.76 -0.49   0.62      0.68       -1.86        1.11
walled garden          -0.33      0.72      0.69 -0.48   0.63      0.67       -1.68        1.01
age                     0.01      1.01      0.01  0.62   0.54      0.90       -0.01        0.03
female                  1.32      3.73      0.43  3.04 <0.005      8.70        0.47        2.17
aware of notification?  0.93      2.53      0.71  1.31   0.19      2.40       -0.46        2.32
right measures          0.89      2.45      0.39  2.30   0.02      5.56        0.13        1.66
home                   -0.20      0.82      0.58 -0.35   0.72      0.46       -1.34        0.93
camera                 -0.34      0.71      0.57 -0.60   0.55      0.86       -1.46        0.78
multiple                0.13      1.14      0.62  0.21   0.83      0.27       -1.08        1.34
unknown                 0.38      1.46      0.69  0.55   0.58      0.78       -0.97        1.72
---
Concordance = 0.66
Log-likelihood ratio test = 23.45 on 11 df, -log2(p)=6.03
Proportional hazard assumption looks okay.
```

Including male:



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -253.94
  time fit was run = 2019-07-25 12:53:59 UTC

---
                        coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
market                 -0.29      0.75      0.35 -0.85 0.40      1.34       -0.97        0.38
email                  -0.27      0.76      0.75 -0.36 0.72      0.48       -1.74        1.20
walled garden          -0.15      0.86      0.68 -0.22 0.83      0.27       -1.48        1.19
age                     0.00      1.00      0.01  0.48 0.63      0.67       -0.02        0.02
male                   -1.03      0.36      0.37 -2.76 0.01      7.43       -1.76       -0.30
aware of notification?  0.85      2.35      0.70  1.22 0.22      2.18       -0.51        2.22
right measures          0.67      1.95      0.36  1.85 0.06      3.96       -0.04        1.37
home                   -0.10      0.90      0.57 -0.18 0.86      0.22       -1.23        1.02
camera                 -0.58      0.56      0.59 -0.98 0.33      1.62       -1.74        0.58
multiple                0.21      1.23      0.61  0.34 0.74      0.44       -0.99        1.41
unknown                 0.39      1.48      0.68  0.58 0.56      0.83       -0.94        1.73
---
Concordance = 0.66
Log-likelihood ratio test = 22.49 on 11 df, -log2(p)=5.59
Proportional hazard assumption looks okay.
```

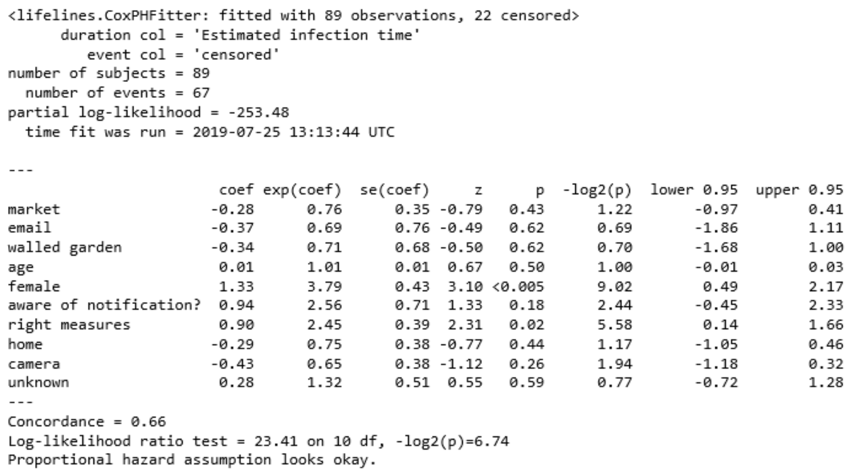
The model including ‘female’ has higher reliable parameters and a better overall fit (the partial log-likelihood is higher for equal degrees of freedom).

In the next steps, the variable ‘male’ is excluded due to the high covariance between this variable and ‘female’. Due to this exclusion, the coding of the dummy variable ‘female’ is changed. 1 = female subscriber, and 0 = male subscriber and subscribers of unknown gender.

Step 1

The ‘multiple’ variable is the least reliable and excluded

Modeling the Cox model for [market, e-mail, walled garden, age, female, aware of notification, right measures, camera, home, unknown]



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -253.48
  time fit was run = 2019-07-25 13:13:44 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                 -0.28      0.76      0.35 -0.79   0.43      1.22       -0.97        0.41
email                  -0.37      0.69      0.76 -0.49   0.62      0.69       -1.86        1.11
walled garden          -0.34      0.71      0.68 -0.50   0.62      0.70       -1.68        1.00
age                     0.01      1.01      0.01  0.67   0.50      1.00       -0.01        0.03
female                  1.33      3.79      0.43  3.10 <0.005      9.02        0.49        2.17
aware of notification?  0.94      2.56      0.71  1.33   0.18      2.44       -0.45        2.33
right measures          0.90      2.45      0.39  2.31   0.02      5.58        0.14        1.66
home                   -0.29      0.75      0.38 -0.77   0.44      1.17       -1.05        0.46
camera                 -0.43      0.65      0.38 -1.12   0.26      1.94       -1.18        0.32
unknown                 0.28      1.32      0.51  0.55   0.59      0.77       -0.72        1.28
---
Concordance = 0.66
Log-likelihood ratio test = 23.41 on 10 df, -log2(p)=6.74
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,02

Log-Likelihood Ratio Statistic (LRS) = 0,04

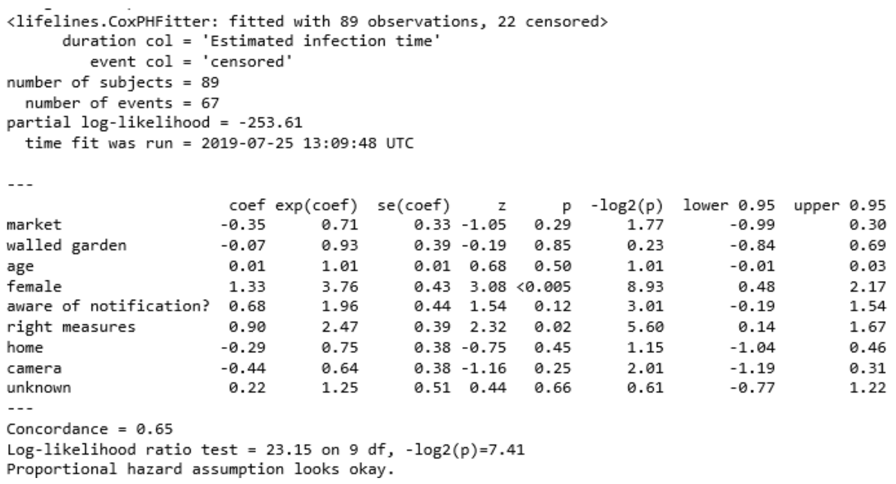
Consulting the chi-square distribution for 0,04 on 1 df: p>0,75

Model 1 is better than model 0

Step 2

The e-mail and walled garden variable are the least reliable. E-mail is excluded because the model of the complete dataset (appendix J.3) has shown that walled garden is a significant covariate and B)

Modeling the Cox model for [market, walled garden, age, female, aware of notification, right measures, home, camera, unknown]



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -253.61
  time fit was run = 2019-07-25 13:09:48 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                 -0.35      0.71      0.33 -1.05   0.29      1.77       -0.99        0.30
walled garden          -0.07      0.93      0.39 -0.19   0.85      0.23       -0.84        0.69
age                     0.01      1.01      0.01  0.68   0.50      1.01       -0.01        0.03
female                  1.33      3.76      0.43  3.08 <0.005      8.93        0.48        2.17
aware of notification?  0.68      1.96      0.44  1.54   0.12      3.01       -0.19        1.54
right measures          0.90      2.47      0.39  2.32   0.02      5.60        0.14        1.67
home                   -0.29      0.75      0.38 -0.75   0.45      1.15       -1.04        0.46
camera                 -0.44      0.64      0.38 -1.16   0.25      2.01       -1.19        0.31
unknown                 0.22      1.25      0.51  0.44   0.66      0.61       -0.77        1.22
---
Concordance = 0.65
Log-likelihood ratio test = 23.15 on 9 df, -log2(p)=7.41
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,13

Log-Likelihood Ratio Statistic (LRS) = 0,26

Consulting the chi-square distribution for 0,26 on 1 df: p>0,50

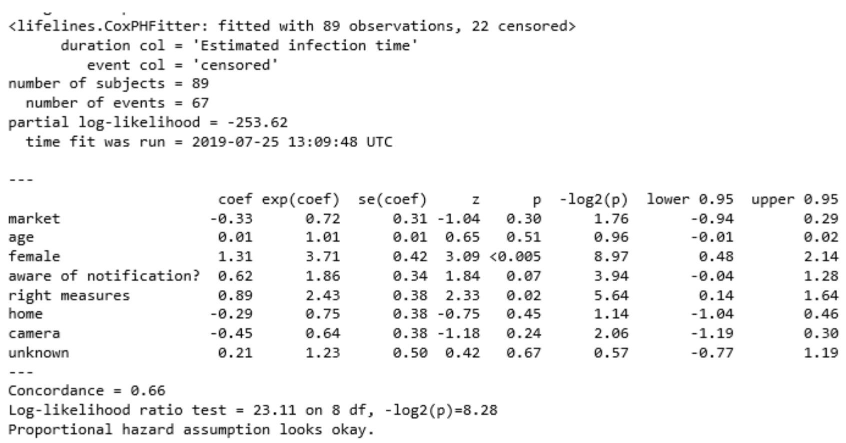
Model 2 is better than model 1

Step 3

The ‘walled garden’ variable is the least reliable and excluded

Modeling the Cox model for [market, age, female, aware of notification, right measures, camera, home, unknown]

<table>
<tr><td></td><td>



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -253.62
  time fit was run = 2019-07-25 13:09:48 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                 -0.33      0.72      0.31 -1.04   0.30      1.76       -0.94        0.29
age                     0.01      1.01      0.01  0.65   0.51      0.96       -0.01        0.02
female                  1.31      3.71      0.42  3.09 <0.005      8.97        0.48        2.14
aware of notification?  0.62      1.86      0.34  1.84   0.07      3.94       -0.04        1.28
right measures          0.89      2.43      0.38  2.33   0.02      5.64        0.14        1.64
home                   -0.29      0.75      0.38 -0.75   0.45      1.14       -1.04        0.46
camera                 -0.45      0.64      0.38 -1.18   0.24      2.06       -1.19        0.30
unknown                 0.21      1.23      0.50  0.42   0.67      0.57       -0.77        1.19
---
Concordance = 0.66
Log-likelihood ratio test = 23.11 on 8 df, -log2(p)=8.28
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,01

Log-Likelihood Ratio Statistic (LRS) = 0,02

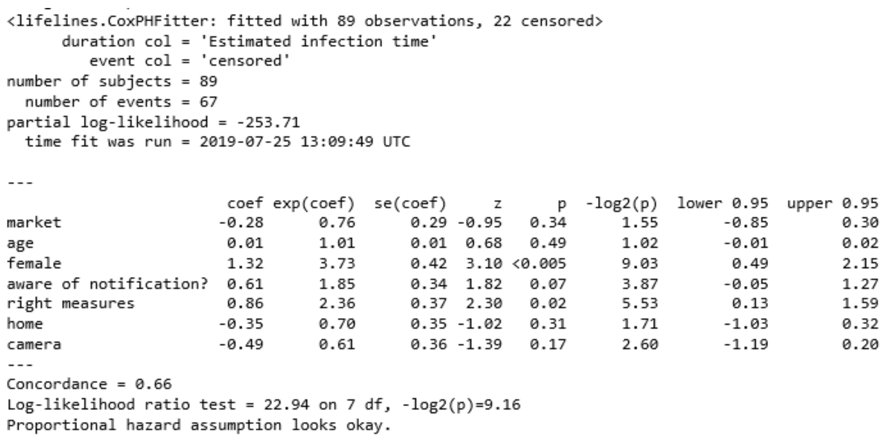
Consulting the chi-square distribution for 0,02 on 1 df: p>0,75

Model 3 is better than model 2

</td></tr>
<tr><td>Step 4</td><td>

The ‘unknown’ variable is the least reliable and excluded

Modeling the Cox model for [market, age, female, aware of notification, right measures, camera, home]



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -253.71
  time fit was run = 2019-07-25 13:09:49 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                 -0.28      0.76      0.29 -0.95   0.34      1.55       -0.85        0.30
age                     0.01      1.01      0.01  0.68   0.49      1.02       -0.01        0.02
female                  1.32      3.73      0.42  3.10 <0.005      9.03        0.49        2.15
aware of notification?  0.61      1.85      0.34  1.82   0.07      3.87       -0.05        1.27
right measures          0.86      2.36      0.37  2.30   0.02      5.53        0.13        1.59
home                   -0.35      0.70      0.35 -1.02   0.31      1.71       -1.03        0.32
camera                 -0.49      0.61      0.36 -1.39   0.17      2.60       -1.19        0.20
---
Concordance = 0.66
Log-likelihood ratio test = 22.94 on 7 df, -log2(p)=9.16
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,09

Log-Likelihood Ratio Statistic (LRS) = 0,18

Consulting the chi-square distribution for 0,18 on 1 df: p>0,50

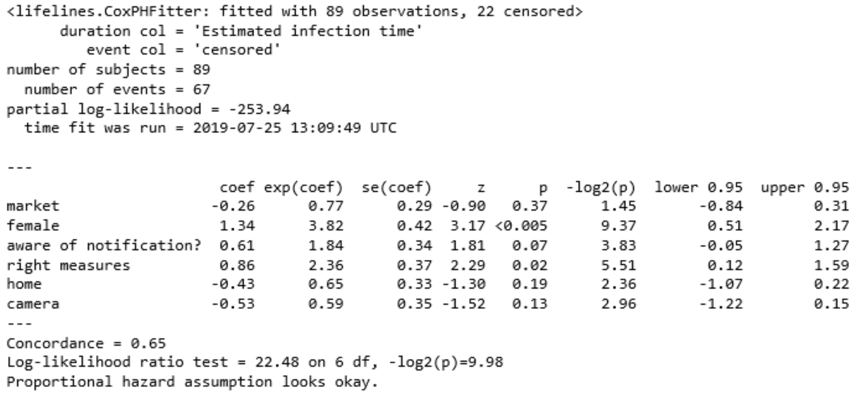
Model 4 is better than model 3

</td></tr>
<tr><td>Step 5</td><td>

The ‘age’ variable is the least reliable and excluded

Modeling the Cox model for [market, female, aware of notification, right measures, camera, home]

</td></tr>
</table>



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -253.94
  time fit was run = 2019-07-25 13:09:49 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                 -0.26      0.77      0.29 -0.90   0.37      1.45       -0.84        0.31
female                  1.34      3.82      0.42  3.17 <0.005      9.37        0.51        2.17
aware of notification?  0.61      1.84      0.34  1.81   0.07      3.83       -0.05        1.27
right measures          0.86      2.36      0.37  2.29   0.02      5.51        0.12        1.59
home                   -0.43      0.65      0.33 -1.30   0.19      2.36       -1.07        0.22
camera                 -0.53      0.59      0.35 -1.52   0.13      2.96       -1.22        0.15
---
Concordance = 0.65
Log-likelihood ratio test = 22.48 on 6 df, -log2(p)=9.98
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,23

Log-Likelihood Ratio Statistic (LRS) = 0,46

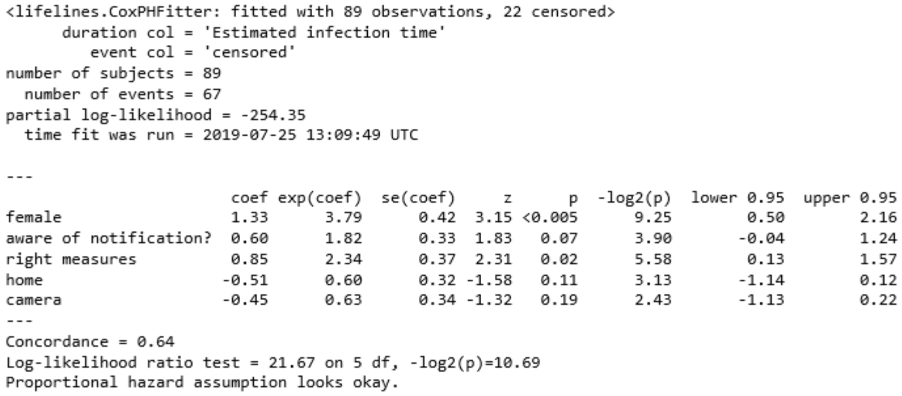
Consulting the chi-square distribution for 0,46 on 1 df: p>0,25

Model 5 is better than model 4

Step 6

The ‘market’ variable is the least reliable and excluded

Modeling the Cox model for [female, aware of notification, right measures, camera, home]



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -254.35
  time fit was run = 2019-07-25 13:09:49 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
female                  1.33      3.79      0.42  3.15 <0.005      9.25        0.50        2.16
aware of notification?  0.60      1.82      0.33  1.83   0.07      3.90       -0.04        1.24
right measures          0.85      2.34      0.37  2.31   0.02      5.58        0.13        1.57
home                   -0.51      0.60      0.32 -1.58   0.11      3.13       -1.14        0.12
camera                 -0.45      0.63      0.34 -1.32   0.19      2.43       -1.13        0.22
---
Concordance = 0.64
Log-likelihood ratio test = 21.67 on 5 df, -log2(p)=10.69
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,41

Log-Likelihood Ratio Statistic (LRS) = 0,82

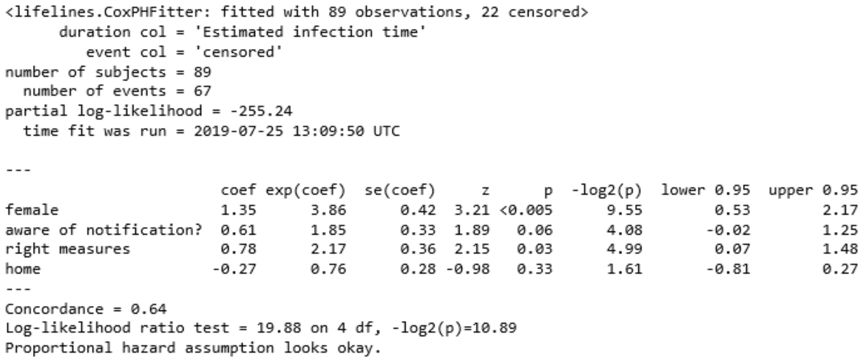
Consulting the chi-square distribution for 0,82 on 1 df: p>0,25

Model 6 is better than model 5

Step 7

The ‘camera’ variable is the least reliable and excluded

Modeling the Cox model for [female, aware of notification, right measures, home]



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -255.24
  time fit was run = 2019-07-25 13:09:50 UTC

---
                        coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
female                  1.35      3.86      0.42  3.21 <0.005      9.55        0.53        2.17
aware of notification?  0.61      1.85      0.33  1.89   0.06      4.08       -0.02        1.25
right measures          0.78      2.17      0.36  2.15   0.03      4.99        0.07        1.48
home                   -0.27      0.76      0.28 -0.98   0.33      1.61       -0.81        0.27
---
Concordance = 0.64
Log-likelihood ratio test = 19.88 on 4 df, -log2(p)=10.89
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,89

Log-Likelihood Ratio Statistic (LRS) = 1,78

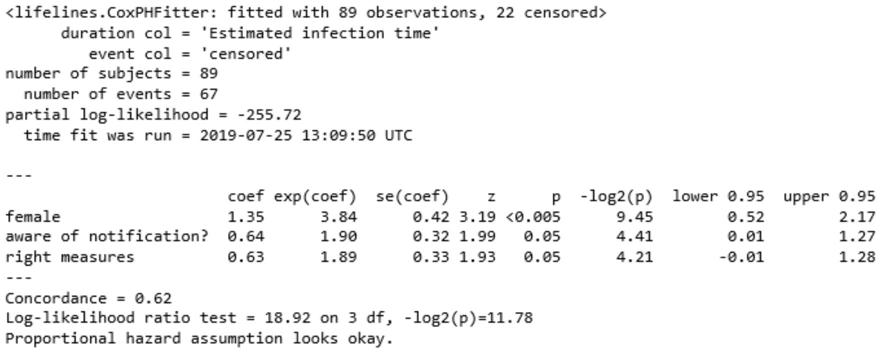
Consulting the chi-square distribution for 1,78on 1 df: p>0,10

Model 7 is better than model 6

Step 8

The 'home' variable is the least reliable and excluded

Modeling the Cox model for [female, aware of notification, right measures]



```
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -255.72
  time fit was run = 2019-07-25 13:09:50 UTC

---
                        coef exp(coef)  se(coef)    z      p  -log2(p)  lower 0.95  upper 0.95
female                  1.35      3.84      0.42 3.19 <0.005      9.45        0.52        2.17
aware of notification?  0.64      1.90      0.32 1.99   0.05      4.41        0.01        1.27
right measures          0.63      1.89      0.33 1.93   0.05      4.21       -0.01        1.28
---
Concordance = 0.62
Log-likelihood ratio test = 18.92 on 3 df, -log2(p)=11.78
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,48

Log-Likelihood Ratio Statistic (LRS) = 0,96

Consulting the chi-square distribution for 0,96 on 1 df: p>0,25

Model 8 is better than model 7

Step 9

All variables are significant. Exclusion of 'aware of notification' or 'right measures' lead to a partial Log-Likelihood of respectively -257,66 and -257,78.

The difference in partial log-likelihood (LL) = -2,06 / -1,94

Log-Likelihood Ratio Statistic (LRS) = 4,12 / 3,88

Consulting the chi-square distribution 4,12 / 3,88 on 1 df, both models: p<0,05

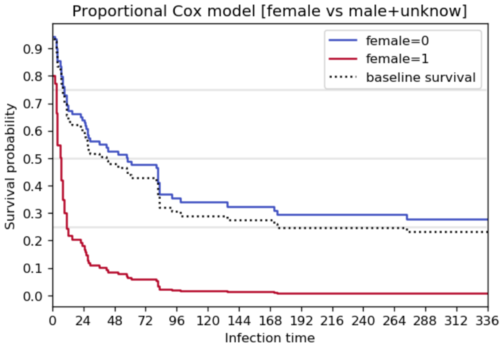
Model 8 is better than a model with less covariates.

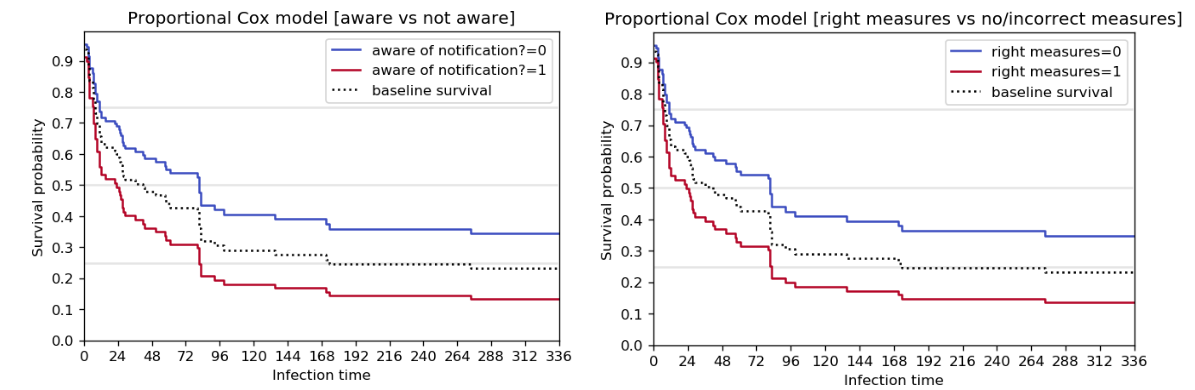
Model 8 is best of all models.

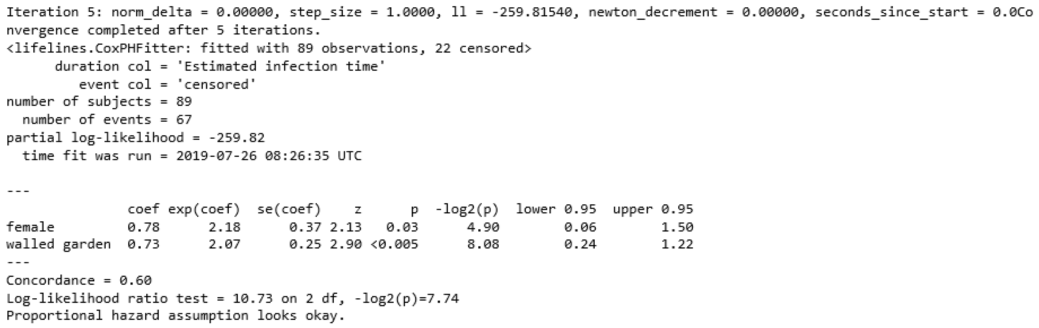
When comparing model 8 with a trivial model, the LRS is 18,92 for 3 degrees of freedom: p<0,01

Model 8 is better than a model without covariates.

Model 8 is accepted







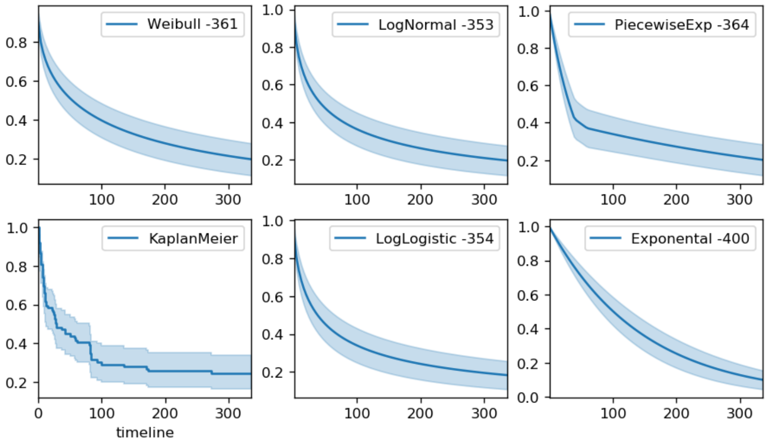
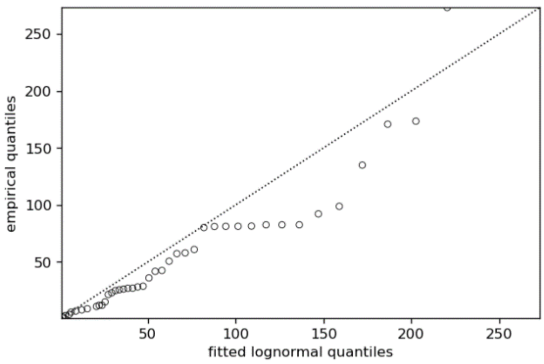
```
Iteration 5: norm_delta = 0.00000, step_size = 1.0000, ll = -259.81540, newton_decrement = 0.00000, seconds_since_start = 0.0Co
nvergence completed after 5 iterations.
<lifelines.CoxPHFitter: fitted with 89 observations, 22 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 89
  number of events = 67
partial log-likelihood = -259.82
  time fit was run = 2019-07-26 08:26:35 UTC

---
               coef exp(coef)  se(coef)    z      p  -log2(p)  lower 0.95  upper 0.95
female         0.78      2.18      0.37 2.13   0.03      4.90        0.06        1.50
walled garden  0.73      2.07      0.25 2.90 <0.005      8.08        0.24        1.22
---
Concordance = 0.60
Log-likelihood ratio test = 10.73 on 2 df, -log2(p)=7.74
Proportional hazard assumption looks okay.
```

AIC model 8: -2 *(-255,72) + (3+3) = 517,44

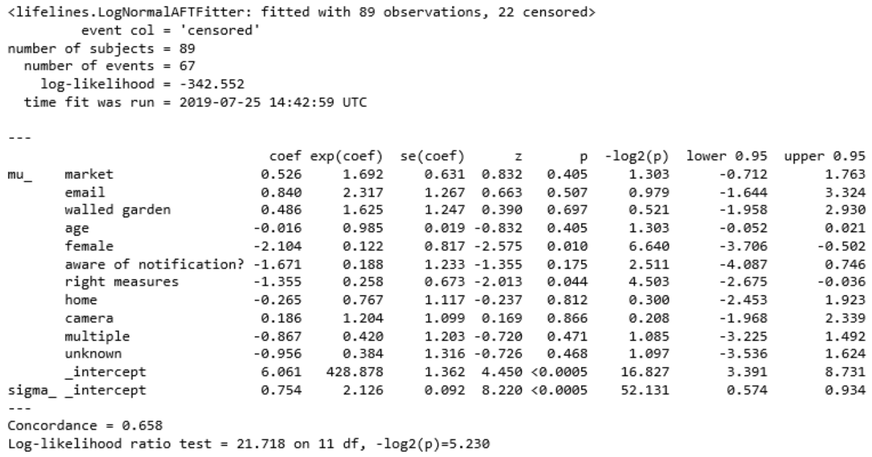
AIC model with walled garden as substitute: -2 * (-259,82) + (2+2) = 521,84

Model 8 has a lower AIC estimate and is thus better than model wherein tha variables awareness and right measures are substituted with the variable walled garden.

K.4 AFT modeling steps

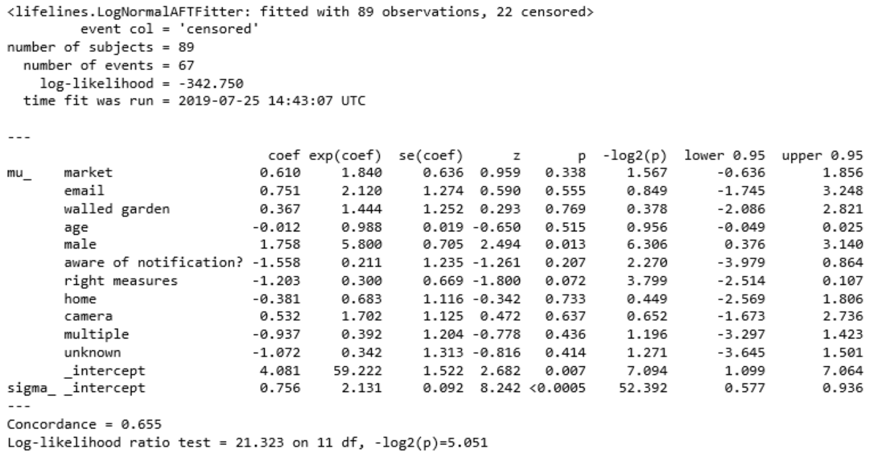
Step 0	Following the similar line of reasoning of appendix J.4, the fitted distributions in figure 57 show that the LogNormal distribution has the best goodness of fit.  Figure 55 Distribution fits for the survival curve of all observations Figure 58 shows the quantile-quantile (Q-Q) plot to compare the fitted LogNormal distribution with the empirical distribution. This Q-Q plot has the similar shape as the Q-Q plot in appendix J.4, but with less dots because the dataset contains less entries. The Q-Q-plot shows that until 81 hours, the data has quite the same shape as the fitted LogNormal distribution. The empirical distribution is a bit more concentrated than the fitted distribution. Then there is a spike of identical values of 81/82 hours (horizontal line of dots). This can also be seen in the Kaplan-Meier plot in chapter 6, which shows a drop around this time. After this spike, the dots form a steep vertical line, which indicates there is a gap in values.  Figure 56 Q-Q plot LogNormal distribution
Step 1	Similar to the Cox modeling in the previous section, we first estimate two models (including female and including male) so we can decide which variable to continue with.

Including female:



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
 number of subjects = 89
   number of events = 67
     log-likelihood = -342.552
   time fit was run = 2019-07-25 14:42:59 UTC

---
                           coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market             0.526     1.692     0.631  0.832   0.405     1.303      -0.712       1.763
       email              0.840     2.317     1.267  0.663   0.507     0.979      -1.644       3.324
       walled garden      0.486     1.625     1.247  0.390   0.697     0.521      -1.958       2.930
       age               -0.016     0.985     0.019 -0.832   0.405     1.303      -0.052       0.021
       female            -2.104     0.122     0.817 -2.575   0.010     6.640      -3.706      -0.502
       aware of notification? -1.671 0.188    1.233 -1.355   0.175     2.511      -4.087       0.746
       right measures    -1.355     0.258     0.673 -2.013   0.044     4.503      -2.675      -0.036
       home              -0.265     0.767     1.117 -0.237   0.812     0.300      -2.453       1.923
       camera             0.186     1.204     1.099  0.169   0.866     0.208      -1.968       2.339
       multiple          -0.867     0.420     1.203 -0.720   0.471     1.085      -3.225       1.492
       unknown           -0.956     0.384     1.316 -0.726   0.468     1.097      -3.536       1.624
       _intercept         6.061   428.878     1.362  4.450 <0.0005    16.827       3.391       8.731
sigma_ _intercept         0.754     2.126     0.092  8.220 <0.0005    52.131       0.574       0.934
---
Concordance = 0.658
Log-likelihood ratio test = 21.718 on 11 df, -log2(p)=5.230
```

Including male:



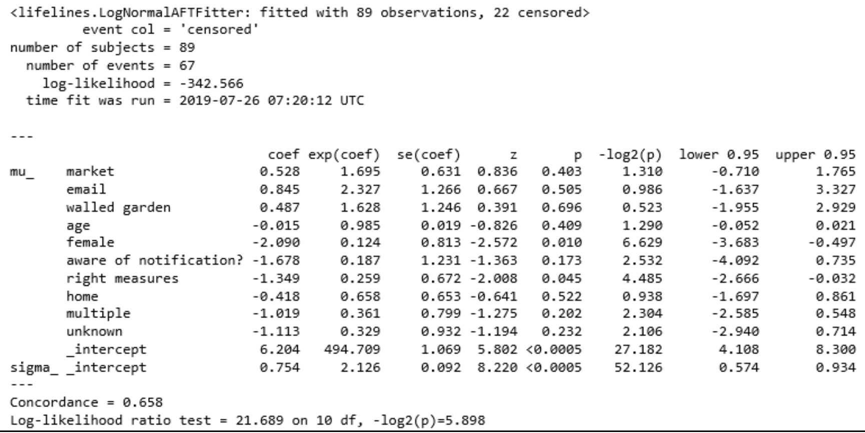
```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
 number of subjects = 89
   number of events = 67
     log-likelihood = -342.750
   time fit was run = 2019-07-25 14:43:07 UTC

---
                           coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market             0.610     1.840     0.636  0.959   0.338     1.567      -0.636       1.856
       email              0.751     2.120     1.274  0.590   0.555     0.849      -1.745       3.248
       walled garden      0.367     1.444     1.252  0.293   0.769     0.378      -2.086       2.821
       age               -0.012     0.988     0.019 -0.650   0.515     0.956      -0.049       0.025
       male               1.758     5.800     0.705  2.494   0.013     6.306       0.376       3.140
       aware of notification? -1.558 0.211    1.235 -1.261   0.207     2.270      -3.979       0.864
       right measures    -1.203     0.300     0.669 -1.800   0.072     3.799      -2.514       0.107
       home              -0.381     0.683     1.116 -0.342   0.733     0.449      -2.569       1.806
       camera             0.532     1.702     1.125  0.472   0.637     0.652      -1.673       2.736
       multiple          -0.937     0.392     1.204 -0.778   0.436     1.196      -3.297       1.423
       unknown           -1.072     0.342     1.313 -0.816   0.414     1.271      -3.645       1.501
       _intercept         4.081    59.222     1.522  2.682   0.007     7.094       1.099       7.064
sigma_ _intercept         0.756     2.131     0.092  8.242 <0.0005    52.392       0.577       0.936
---
Concordance = 0.655
Log-likelihood ratio test = 21.323 on 11 df, -log2(p)=5.051
```

The model including ‘female’ has a better overall fit (the log-likelihood is higher for equal degrees of freedom). Similar to the Cox model, the next steps will include ‘female’. (Dummy coding: 1 = female subscriber, and 0 = male subscriber and subscribers of unknown gender)

Step 2

The camera variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, e-mail, walled garden, age, female, aware of notification, right measures, home, multiple, unknown]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
 number of subjects = 89
   number of events = 67
     log-likelihood = -342.566
   time fit was run = 2019-07-26 07:20:12 UTC

---
                           coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market             0.528     1.695     0.631  0.836   0.403     1.310      -0.710       1.765
       email              0.845     2.327     1.266  0.667   0.505     0.986      -1.637       3.327
       walled garden      0.487     1.628     1.246  0.391   0.696     0.523      -1.955       2.929
       age               -0.015     0.985     0.019 -0.826   0.409     1.290      -0.052       0.021
       female            -2.090     0.124     0.813 -2.572   0.010     6.629      -3.683      -0.497
       aware of notification? -1.678 0.187    1.231 -1.363   0.173     2.532      -4.092       0.735
       right measures    -1.349     0.259     0.672 -2.008   0.045     4.485      -2.666      -0.032
       home              -0.418     0.658     0.653 -0.641   0.522     0.938      -1.697       0.861
       multiple          -1.019     0.361     0.799 -1.275   0.202     2.304      -2.585       0.548
       unknown           -1.113     0.329     0.932 -1.194   0.232     2.106      -2.940       0.714
       _intercept         6.204   494.709     1.069  5.802 <0.0005    27.182       4.108       8.300
sigma_ _intercept         0.754     2.126     0.092  8.220 <0.0005    52.126       0.574       0.934
---
Concordance = 0.658
Log-likelihood ratio test = 21.689 on 10 df, -log2(p)=5.898
```

The difference in partial log-likelihood (LL) = -0,014

Log-Likelihood Ratio Statistic (LRS) = 0,028

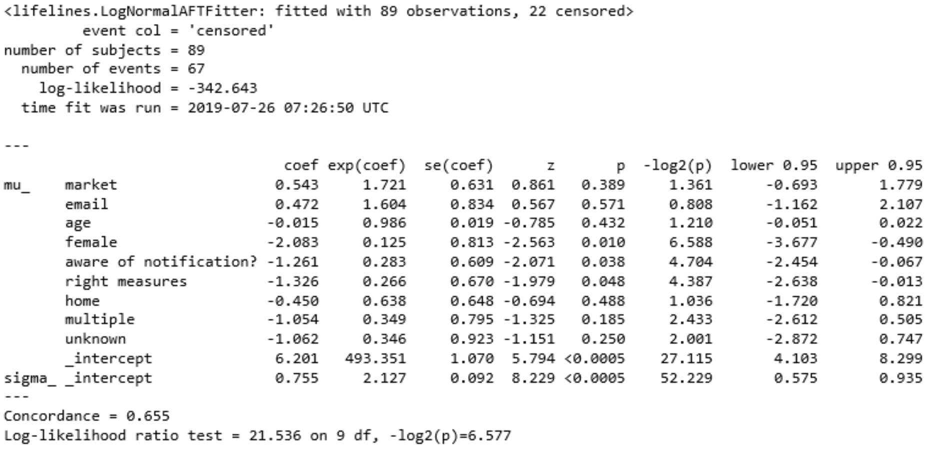
Consulting the chi-square distribution for 0,028 on 1 df: p>0,75

Model 2 is better than model 1

Step3

The walled garden variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, e-mail, age, female, aware of notification, right measures, home, multiple, unknown]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -342.643
  time fit was run = 2019-07-26 07:26:50 UTC

---
                             coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market               0.543     1.721     0.631  0.861   0.389     1.361      -0.693       1.779
       email                0.472     1.604     0.834  0.567   0.571     0.808      -1.162       2.107
       age                 -0.015     0.986     0.019 -0.785   0.432     1.210      -0.051       0.022
       female              -2.083     0.125     0.813 -2.563   0.010     6.588      -3.677      -0.490
       aware of notification? -1.261     0.283     0.609 -2.071   0.038     4.704      -2.454      -0.067
       right measures      -1.326     0.266     0.670 -1.979   0.048     4.387      -2.638      -0.013
       home                -0.450     0.638     0.648 -0.694   0.488     1.036      -1.720       0.821
       multiple            -1.054     0.349     0.795 -1.325   0.185     2.433      -2.612       0.505
       unknown             -1.062     0.346     0.923 -1.151   0.250     2.001      -2.872       0.747
       _intercept           6.201   493.351     1.070  5.794 <0.0005    27.115       4.103       8.299
sigma_ _intercept           0.755     2.127     0.092  8.229 <0.0005    52.229       0.575       0.935
---
Concordance = 0.655
Log-likelihood ratio test = 21.536 on 9 df, -log2(p)=6.577
```

The difference in partial log-likelihood (LL) = -0,077

Log-Likelihood Ratio Statistic (LRS) = 0,154

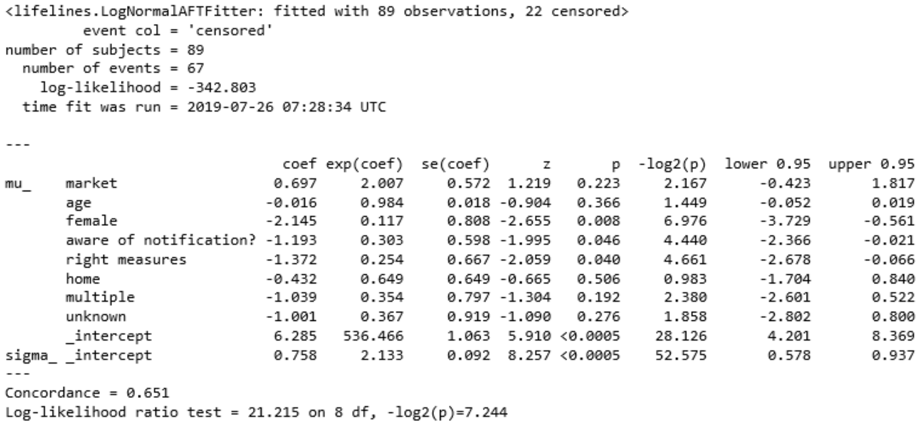
Consulting the chi-square distribution for 0,154 on 1 df: p>0,50

Model 3 is better than model 2

Step 4

The e-mail variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, age, female, aware of notification, right measures, home, multiple, unknown]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -342.803
  time fit was run = 2019-07-26 07:28:34 UTC

---
                             coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market               0.697     2.007     0.572  1.219   0.223     2.167      -0.423       1.817
       age                 -0.016     0.984     0.018 -0.904   0.366     1.449      -0.052       0.019
       female              -2.145     0.117     0.808 -2.655   0.008     6.976      -3.729      -0.561
       aware of notification? -1.193     0.303     0.598 -1.995   0.046     4.440      -2.366      -0.021
       right measures      -1.372     0.254     0.667 -2.059   0.040     4.661      -2.678      -0.066
       home                -0.432     0.649     0.649 -0.665   0.506     0.983      -1.704       0.840
       multiple            -1.039     0.354     0.797 -1.304   0.192     2.380      -2.601       0.522
       unknown             -1.001     0.367     0.919 -1.090   0.276     1.858      -2.802       0.800
       _intercept           6.285   536.466     1.063  5.910 <0.0005    28.126       4.201       8.369
sigma_ _intercept           0.758     2.133     0.092  8.257 <0.0005    52.575       0.578       0.937
---
Concordance = 0.651
Log-likelihood ratio test = 21.215 on 8 df, -log2(p)=7.244
```

The difference in partial log-likelihood (LL) = -0,16

Log-Likelihood Ratio Statistic (LRS) = 0,32

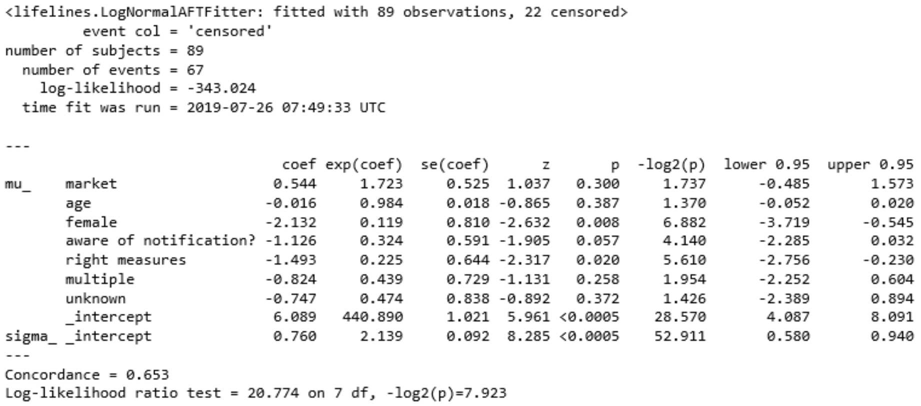
Consulting the chi-square distribution for 0,32 on 1 df: p>0,50

Model 4 is better than model 3

Step 5

The home variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, age, female, aware of notification, right measures, multiple, unknown]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -343.024
  time fit was run = 2019-07-26 07:49:33 UTC

---
                                coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_     market                 0.544     1.723     0.525  1.037   0.300     1.737      -0.485       1.573
        age                   -0.016     0.984     0.018 -0.865   0.387     1.370      -0.052       0.020
        female                -2.132     0.119     0.810 -2.632   0.008     6.882      -3.719      -0.545
        aware of notification? -1.126    0.324     0.591 -1.905   0.057     4.140      -2.285       0.032
        right measures        -1.493     0.225     0.644 -2.317   0.020     5.610      -2.756      -0.230
        multiple              -0.824     0.439     0.729 -1.131   0.258     1.954      -2.252       0.604
        unknown               -0.747     0.474     0.838 -0.892   0.372     1.426      -2.389       0.894
        _intercept             6.089   440.890     1.021  5.961 <0.0005    28.570       4.087       8.091
sigma_ _intercept              0.760     2.139     0.092  8.285 <0.0005    52.911       0.580       0.940
---
Concordance = 0.653
Log-likelihood ratio test = 20.774 on 7 df, -log2(p)=7.923
```

The difference in partial log-likelihood (LL) = -0,221

Log-Likelihood Ratio Statistic (LRS) = 0,442

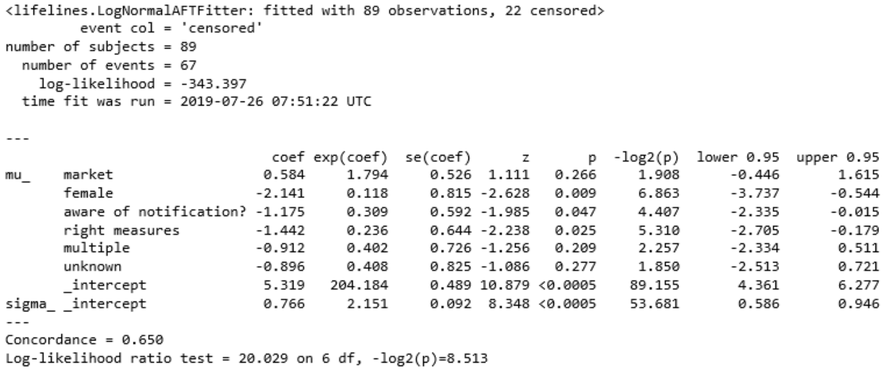
Consulting the chi-square distribution for 0,442 on 1 df: p>0,50

Model 5 is better than model 4

Step 6

The age variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, female, aware of notification, right measures, multiple, unknown]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -343.397
  time fit was run = 2019-07-26 07:51:22 UTC

---
                                coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_     market                 0.584     1.794     0.526  1.111   0.266     1.908      -0.446       1.615
        female                -2.141     0.118     0.815 -2.628   0.009     6.863      -3.737      -0.544
        aware of notification? -1.175    0.309     0.592 -1.985   0.047     4.407      -2.335      -0.015
        right measures        -1.442     0.236     0.644 -2.238   0.025     5.310      -2.705      -0.179
        multiple              -0.912     0.402     0.726 -1.256   0.209     2.257      -2.334       0.511
        unknown               -0.896     0.408     0.825 -1.086   0.277     1.850      -2.513       0.721
        _intercept             5.319   204.184     0.489 10.879 <0.0005    89.155       4.361       6.277
sigma_ _intercept              0.766     2.151     0.092  8.348 <0.0005    53.681       0.586       0.946
---
Concordance = 0.650
Log-likelihood ratio test = 20.029 on 6 df, -log2(p)=8.513
```

The difference in partial log-likelihood (LL) = -0,373

Log-Likelihood Ratio Statistic (LRS) = 0,746

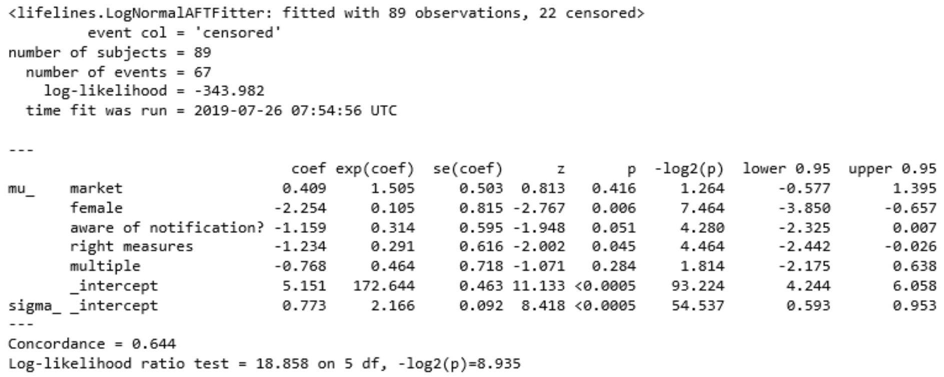
Consulting the chi-square distribution for 0,746 on 1 df: p>0,25

Model 6 is better than model 5

Step 7

The unknown variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, female, aware of notification, right measures, multiple]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -343.982
  time fit was run = 2019-07-26 07:54:56 UTC

---
                                coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                  0.409     1.505     0.503  0.813   0.416     1.264      -0.577       1.395
       female                 -2.254     0.105     0.815 -2.767   0.006     7.464      -3.850      -0.657
       aware of notification? -1.159     0.314     0.595 -1.948   0.051     4.280      -2.325       0.007
       right measures         -1.234     0.291     0.616 -2.002   0.045     4.464      -2.442      -0.026
       multiple               -0.768     0.464     0.718 -1.071   0.284     1.814      -2.175       0.638
       _intercept              5.151   172.644     0.463 11.133 <0.0005    93.224       4.244       6.058
sigma_ _intercept              0.773     2.166     0.092  8.418 <0.0005    54.537       0.593       0.953
---
Concordance = 0.644
Log-likelihood ratio test = 18.858 on 5 df, -log2(p)=8.935
```

The difference in partial log-likelihood (LL) = -0,585

Log-Likelihood Ratio Statistic (LRS) = 1,17

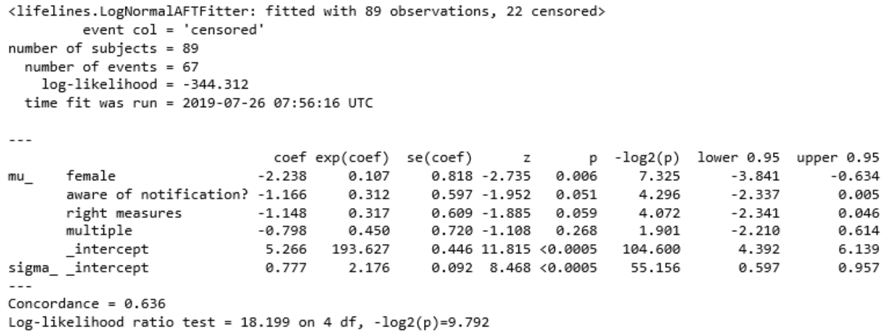
Consulting the chi-square distribution for 1,17 on 1 df: p>0,25

Model 7 is better than model 6

Step 8

The market variable is the least reliable and excluded

Modeling the AFT LogNormal model for [female, aware of notification, right measures, multiple]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -344.312
  time fit was run = 2019-07-26 07:56:16 UTC

---
                                coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    female                 -2.238     0.107     0.818 -2.735   0.006     7.325      -3.841      -0.634
       aware of notification? -1.166     0.312     0.597 -1.952   0.051     4.296      -2.337       0.005
       right measures         -1.148     0.317     0.609 -1.885   0.059     4.072      -2.341       0.046
       multiple               -0.798     0.450     0.720 -1.108   0.268     1.901      -2.210       0.614
       _intercept              5.266   193.627     0.446 11.815 <0.0005   104.600       4.392       6.139
sigma_ _intercept              0.777     2.176     0.092  8.468 <0.0005    55.156       0.597       0.957
---
Concordance = 0.636
Log-likelihood ratio test = 18.199 on 4 df, -log2(p)=9.792
```

The difference in partial log-likelihood (LL) = -0,33

Log-Likelihood Ratio Statistic (LRS) = 0,66

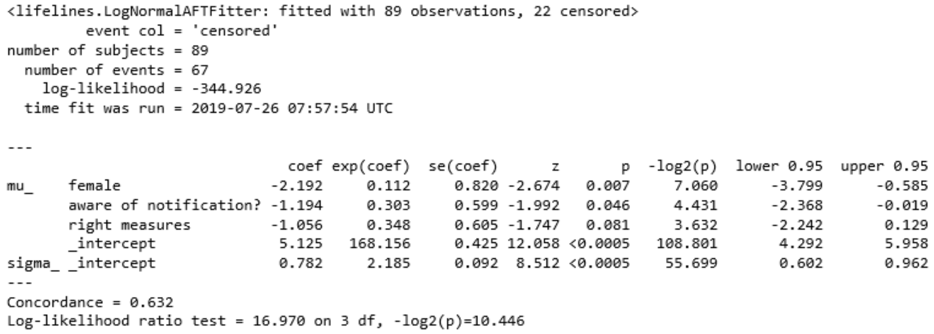
Consulting the chi-square distribution for 0,66 on 1 df: p>0,25

Model 8 is better than model 7

Step 9

The multiple variable is the least reliable and excluded

Modeling the AFT LogNormal model for [female, aware of notification, right measures]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -344.926
  time fit was run = 2019-07-26 07:57:54 UTC

---
                                coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    female                 -2.192     0.112     0.820 -2.674   0.007     7.060      -3.799      -0.585
       aware of notification? -1.194     0.303     0.599 -1.992   0.046     4.431      -2.368      -0.019
       right measures         -1.056     0.348     0.605 -1.747   0.081     3.632      -2.242       0.129
       _intercept              5.125   168.156     0.425 12.058 <0.0005   108.801       4.292       5.958
sigma_ _intercept              0.782     2.185     0.092  8.512 <0.0005    55.699       0.602       0.962
---
Concordance = 0.632
Log-likelihood ratio test = 16.970 on 3 df, -log2(p)=10.446
```

The difference in partial log-likelihood (LL) = -0,614

Log-Likelihood Ratio Statistic (LRS) = 1,228

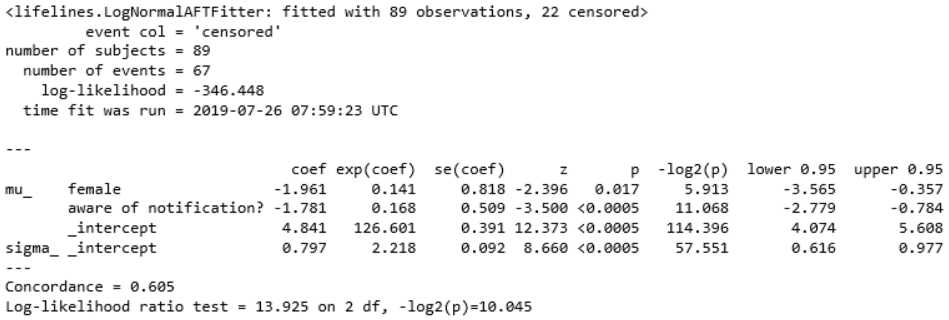
Consulting the chi-square distribution for 1,228 on 1 df: p>0,25

Model 9 is better than model 8

Step 10

The female measures variable is the least reliable and excluded

Modeling the AFT LogNormal model for [aware of notification]



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -346.448
  time fit was run = 2019-07-26 07:59:23 UTC

---
                             coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    female              -1.961     0.141     0.818 -2.396   0.017     5.913      -3.565      -0.357
       aware of notification? -1.781  0.168     0.509 -3.500 <0.0005    11.068      -2.779      -0.784
       _intercept           4.841   126.601     0.391 12.373 <0.0005   114.396       4.074       5.608
sigma_ _intercept           0.797     2.218     0.092  8.660 <0.0005    57.551       0.616       0.977
---
Concordance = 0.605
Log-likelihood ratio test = 13.925 on 2 df, -log2(p)=10.045
```

The difference in partial log-likelihood (LL) = -1,522

Log-Likelihood Ratio Statistic (LRS) = 3,044

Consulting the chi-square distribution for 3,044 on 1 df: p=0,08

Model 10 is better than model 9

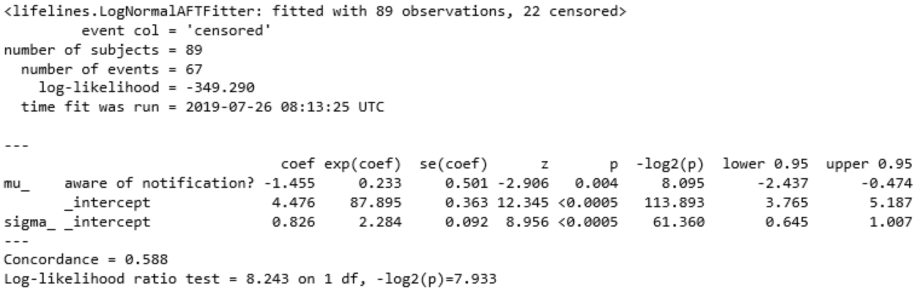
Step 11

All variables are significant. Exclusion of ‘female’ leads to a Log-Likelihood of -349,290

The difference in partial log-likelihood (LL) = -2,842

Log-Likelihood Ratio Statistic (LRS) = 5,684

Consulting the chi-square distribution 5,684 on 1 df: p<0,05



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -349.290
  time fit was run = 2019-07-26 08:13:25 UTC

---
                             coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    aware of notification? -1.455  0.233     0.501 -2.906   0.004     8.095      -2.437      -0.474
       _intercept           4.476    87.895     0.363 12.345 <0.0005   113.893       3.765       5.187
sigma_ _intercept           0.826     2.284     0.092  8.956 <0.0005    61.360       0.645       1.007
---
Concordance = 0.588
Log-likelihood ratio test = 8.243 on 1 df, -log2(p)=7.933
```

Model 10 is better than model 11

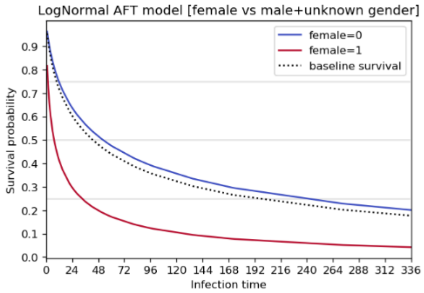
Step 12

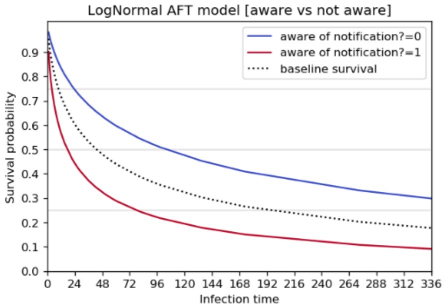
Model 10 is best of all models.

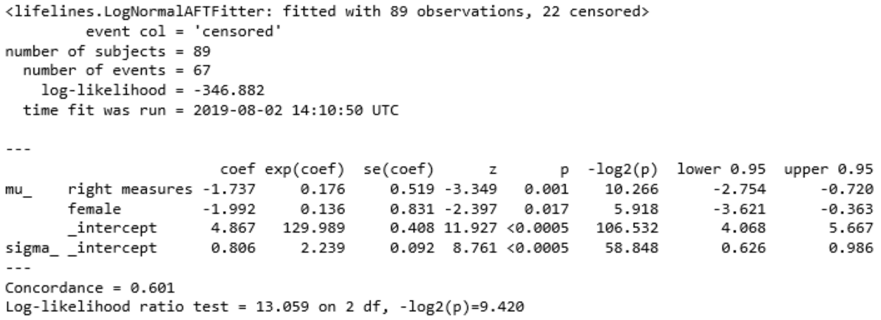
When comparing model 10 with a trivial model, the LRS is 13,925 for 2 degrees of freedom: p<0,01

Model 10 is better than a model without covariates.

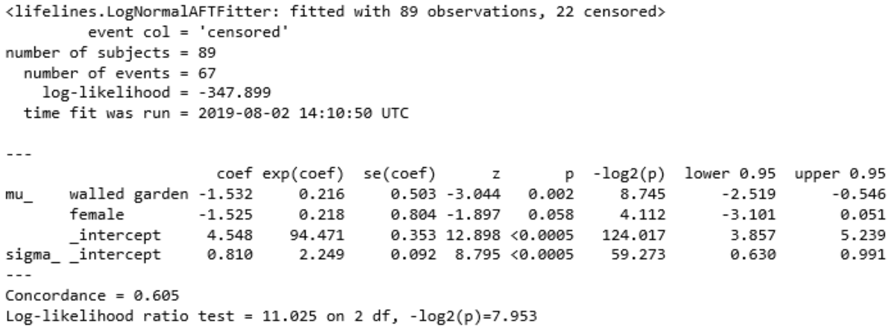
Model 10 is accepted







```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -346.882
  time fit was run = 2019-08-02 14:10:50 UTC

---
                        coef  exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    right measures -1.737      0.176     0.519 -3.349   0.001    10.266      -2.754      -0.720
       female         -1.992      0.136     0.831 -2.397   0.017     5.918      -3.621      -0.363
       _intercept      4.867    129.989     0.408 11.927 <0.0005   106.532       4.068       5.667
sigma_ _intercept      0.806      2.239     0.092  8.761 <0.0005    58.848       0.626       0.986
---
Concordance = 0.601
Log-likelihood ratio test = 13.059 on 2 df, -log2(p)=9.420
```



```
<lifelines.LogNormalAFTFitter: fitted with 89 observations, 22 censored>
         event col = 'censored'
number of subjects = 89
  number of events = 67
    log-likelihood = -347.899
  time fit was run = 2019-08-02 14:10:50 UTC

---
                        coef  exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    walled garden -1.532      0.216     0.503 -3.044   0.002     8.745      -2.519      -0.546
       female        -1.525      0.218     0.804 -1.897   0.058     4.112      -3.101       0.051
       _intercept     4.548     94.471     0.353 12.898 <0.0005   124.017       3.857       5.239
sigma_ _intercept     0.810      2.249     0.092  8.795 <0.0005    59.273       0.630       0.991
---
Concordance = 0.605
Log-likelihood ratio test = 11.025 on 2 df, -log2(p)=7.953
```

Model 10 has a Log-Likelihood of -346,448, which is higher than these models with substituted variables for awareness. Model 10 is thus the best model.

L Modeling step 3 [interviewed, notified consumers]

L.1 Introduction

Of the 89 interviewed consumers that are included in the modeling process (as described in appendix K), 49 consumers were notified. 37 consumers of these were placed in a walled garden and 12 consumers were sent only an e-mail. 9 entries of the 49 are censored.

Using this dataset enables us to use more information obtained from the interviews that concern only notified consumers. In addition to the variables addressed in appendix K.1, this modeling process includes four more dummy variables. Two variables are two stages of the theoretical framework: comprehension and intention. (Strict) compliance not included because we want to know the influence of effective remediation measures, which is already covered by the variable 'right measures'. The two dummy variables are defined and coded as the following:

Understood notification (comprehension): whether the interviewed consumer understood the content of the notification

	Variable 'understood notification?':
The notification was unclear / had some unclear parts	0
The notification was clear to the consumer	1

Intended to comply (intention): whether the interviewed consumer had the intention to take comply with the recommended actions.

	Variable 'intended to comply?:
No intention to comply	0
Intention to comply	1

Consumers were also asked about their experience with the notification. The information about their satisfaction is included using two dummy variables, coded as follows:

	Variable 'satisfied with service'	Variable 'dissatisfied with service'
Satisfied	1	0
Dissatisfied	0	1
Neutral	0	0

Due to the absence of the control group, the dummy variables for the notification mechanisms is changed. The variable 'e-mail' is excluded and the zero-value of 'walled garden' now refers to e-mail notified consumers only.

L.2 Overview data

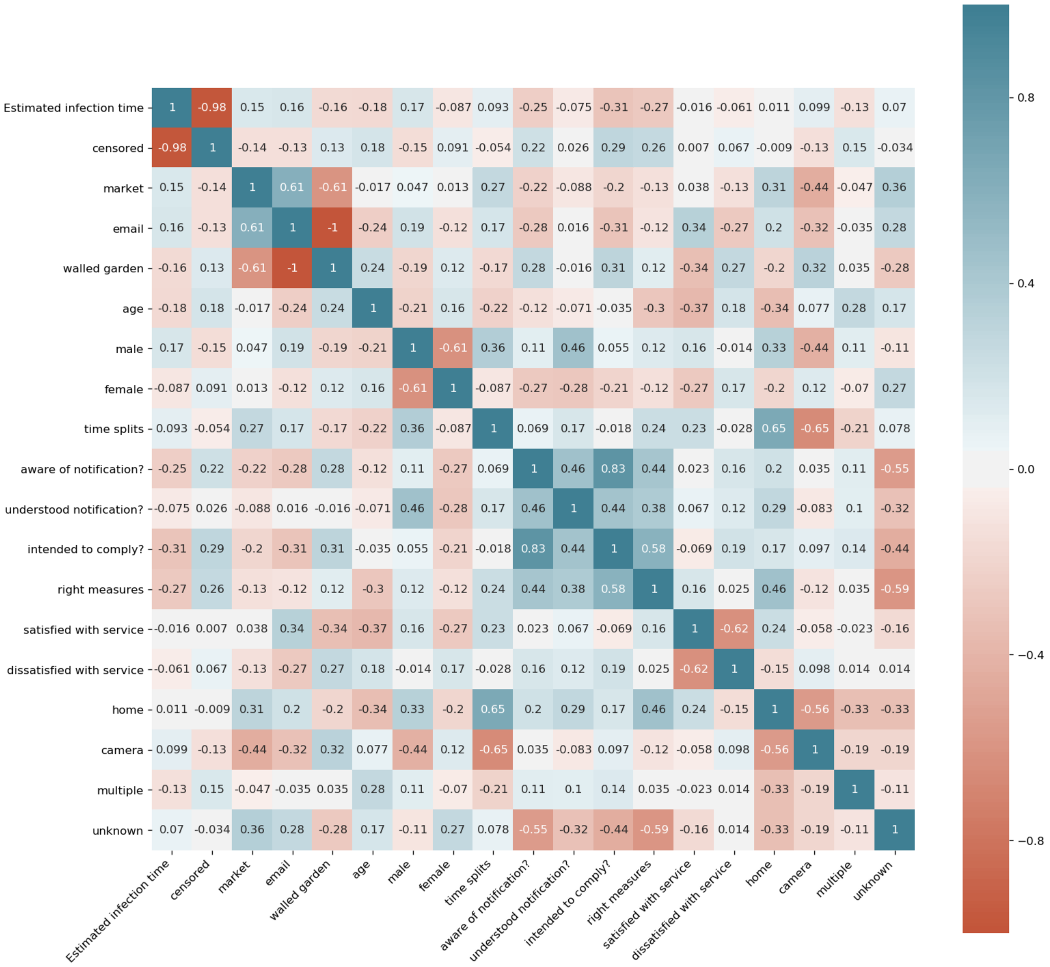
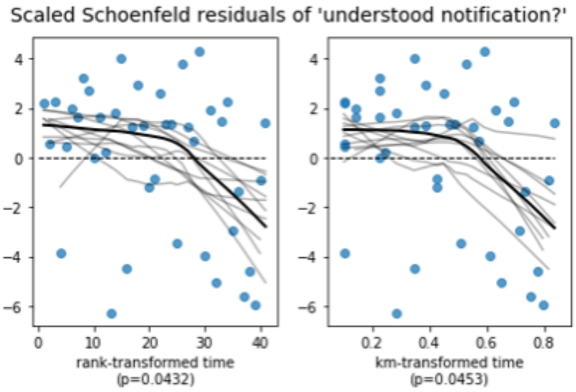


Figure 57 Correlation potential remediation drivers (interviewed, notified consumers)

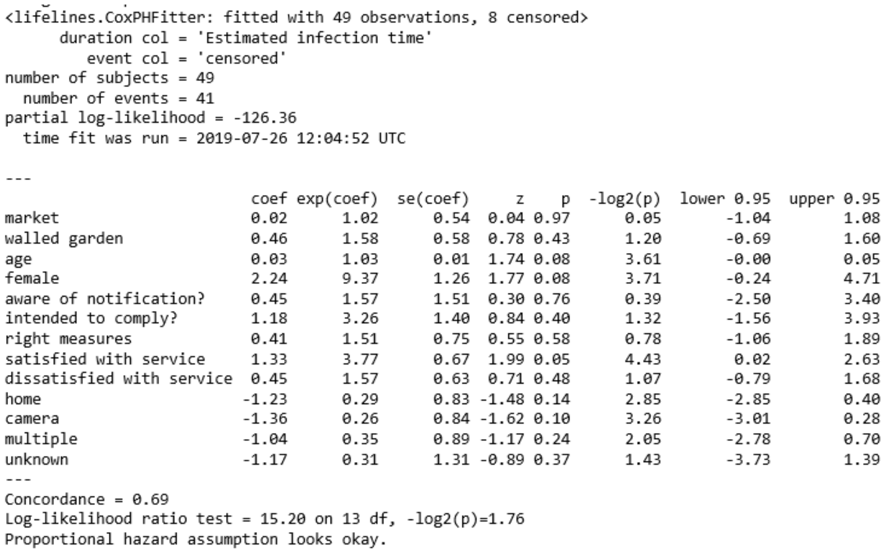
L.3 Cox modeling steps

Step 0	For the same reasons as addressed in appendix J.2, we first check if ‘female’ or ‘male’ is a better covariate to determine which of the two to include. Including ‘male’ leads to a violation of the proportional hazard assumption of the ‘understood notification?’ variable. Complimentary to the statistical test, we can review this violation in this figure (lines should be constant):



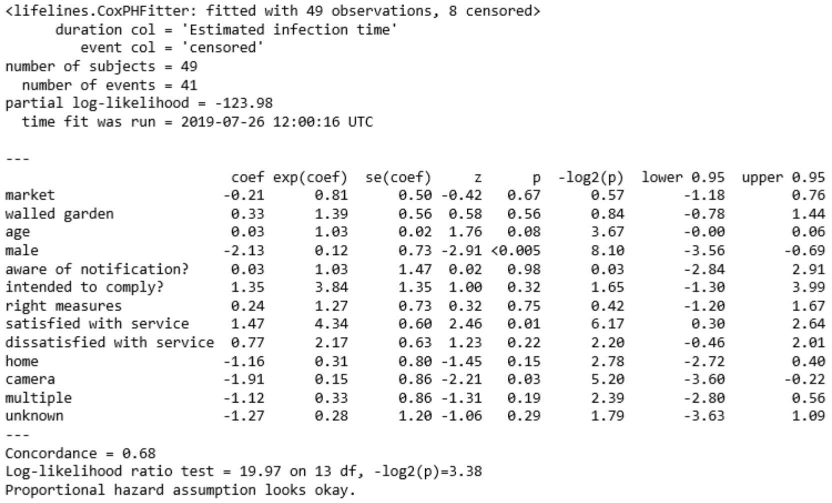
To compare the model including ‘female’ with a model including ‘male’, we delete the variable in question.

Including female:



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -126.36
  time fit was run = 2019-07-26 12:04:52 UTC

---
                           coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
market                     0.02      1.02      0.54  0.04 0.97      0.05       -1.04        1.08
walled garden              0.46      1.58      0.58  0.78 0.43      1.20       -0.69        1.60
age                        0.03      1.03      0.01  1.74 0.08      3.61       -0.00        0.05
female                     2.24      9.37      1.26  1.77 0.08      3.71       -0.24        4.71
aware of notification?     0.45      1.57      1.51  0.30 0.76      0.39       -2.50        3.40
intended to comply?        1.18      3.26      1.40  0.84 0.40      1.32       -1.56        3.93
right measures             0.41      1.51      0.75  0.55 0.58      0.78       -1.06        1.89
satisfied with service     1.33      3.77      0.67  1.99 0.05      4.43        0.02        2.63
dissatisfied with service  0.45      1.57      0.63  0.71 0.48      1.07       -0.79        1.68
home                      -1.23      0.29      0.83 -1.48 0.14      2.85       -2.85        0.40
camera                    -1.36      0.26      0.84 -1.62 0.10      3.26       -3.01        0.28
multiple                  -1.04      0.35      0.89 -1.17 0.24      2.05       -2.78        0.70
unknown                   -1.17      0.31      1.31 -0.89 0.37      1.43       -3.73        1.39
---
Concordance = 0.69
Log-likelihood ratio test = 15.20 on 13 df, -log2(p)=1.76
Proportional hazard assumption looks okay.
```

Including male:



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -123.98
  time fit was run = 2019-07-26 12:00:16 UTC

---
                           coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                    -0.21      0.81      0.50 -0.42   0.67      0.57       -1.18        0.76
walled garden              0.33      1.39      0.56  0.58   0.56      0.84       -0.78        1.44
age                        0.03      1.03      0.02  1.76   0.08      3.67       -0.00        0.06
male                      -2.13      0.12      0.73 -2.91 <0.005      8.10       -3.56       -0.69
aware of notification?     0.03      1.03      1.47  0.02   0.98      0.03       -2.84        2.91
intended to comply?        1.35      3.84      1.35  1.00   0.32      1.65       -1.30        3.99
right measures             0.24      1.27      0.73  0.32   0.75      0.42       -1.20        1.67
satisfied with service     1.47      4.34      0.60  2.46   0.01      6.17        0.30        2.64
dissatisfied with service  0.77      2.17      0.63  1.23   0.22      2.20       -0.46        2.01
home                      -1.16      0.31      0.80 -1.45   0.15      2.78       -2.72        0.40
camera                    -1.91      0.15      0.86 -2.21   0.03      5.20       -3.60       -0.22
multiple                  -1.12      0.33      0.86 -1.31   0.19      2.39       -2.80        0.56
unknown                   -1.27      0.28      1.20 -1.06   0.29      1.79       -3.63        1.09
---
Concordance = 0.68
Log-likelihood ratio test = 19.97 on 13 df, -log2(p)=3.38
Proportional hazard assumption looks okay.
```

The model including ‘male’ has higher reliable parameters and a better overall fit (the partial log-likelihood is higher for equal degrees of freedom).

<table>
<tr>
<td></td>
<td>

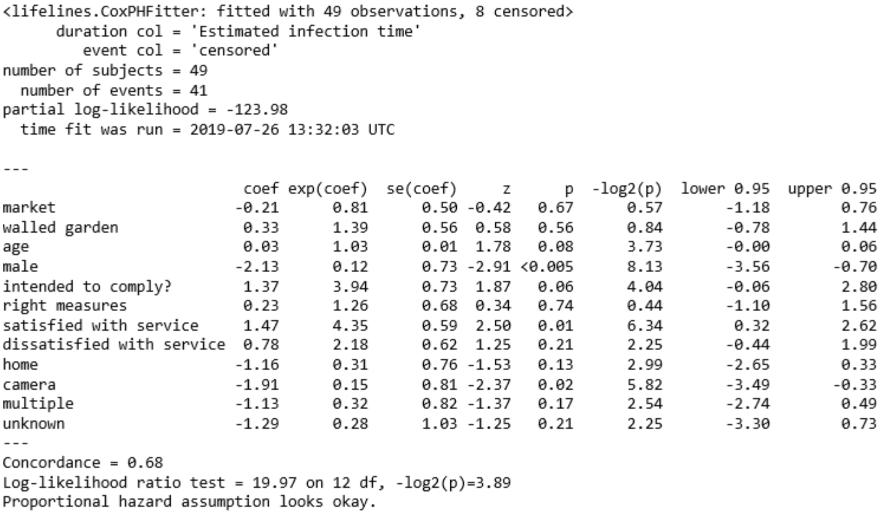
In the next steps, the variable ‘female’ is excluded. Due to this exclusion, the coding of the dummy variable ‘male’ is changed. 1 = male subscriber, and 0 = female subscriber and subscribers of unknown gender.

</td>
</tr>
<tr>
<td>Step 1</td>
<td>

Due to the violation of the proportional hazard assumption, we can either stratify or exclude the ‘understood notification’ variable. We choose for the latter option due to A) the low reliability of the variable when including ‘female’ instead of ‘male’ and B) the high covariance with other variables which may indicate an overlap with the other variables. In the last step of this modeling process, we will include this variable again to check whether we made a misjudgment.

We continue with model 0. The ‘aware of notification’ variable is the least reliable and excluded

Modeling the Cox model for [market, walled garden, age, male, intended to comply, right measures, satisfied with service, dissatisfied with service, home, camera, multiple, unknown]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -123.98
  time fit was run = 2019-07-26 13:32:03 UTC

---
                           coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                    -0.21      0.81      0.50 -0.42   0.67      0.57       -1.18        0.76
walled garden              0.33      1.39      0.56  0.58   0.56      0.84       -0.78        1.44
age                        0.03      1.03      0.01  1.78   0.08      3.73       -0.00        0.06
male                      -2.13      0.12      0.73 -2.91 <0.005      8.13       -3.56       -0.70
intended to comply?        1.37      3.94      0.73  1.87   0.06      4.04       -0.06        2.80
right measures             0.23      1.26      0.68  0.34   0.74      0.44       -1.10        1.56
satisfied with service     1.47      4.35      0.59  2.50   0.01      6.34        0.32        2.62
dissatisfied with service  0.78      2.18      0.62  1.25   0.21      2.25       -0.44        1.99
home                      -1.16      0.31      0.76 -1.53   0.13      2.99       -2.65        0.33
camera                    -1.91      0.15      0.81 -2.37   0.02      5.82       -3.49       -0.33
multiple                  -1.13      0.32      0.82 -1.37   0.17      2.54       -2.74        0.49
unknown                   -1.29      0.28      1.03 -1.25   0.21      2.25       -3.30        0.73
---
Concordance = 0.68
Log-likelihood ratio test = 19.97 on 12 df, -log2(p)=3.89
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,00

Log-Likelihood Ratio Statistic (LRS) = 0,00

Consulting the chi-square distribution for 0,00 on 1 df: p>0,99

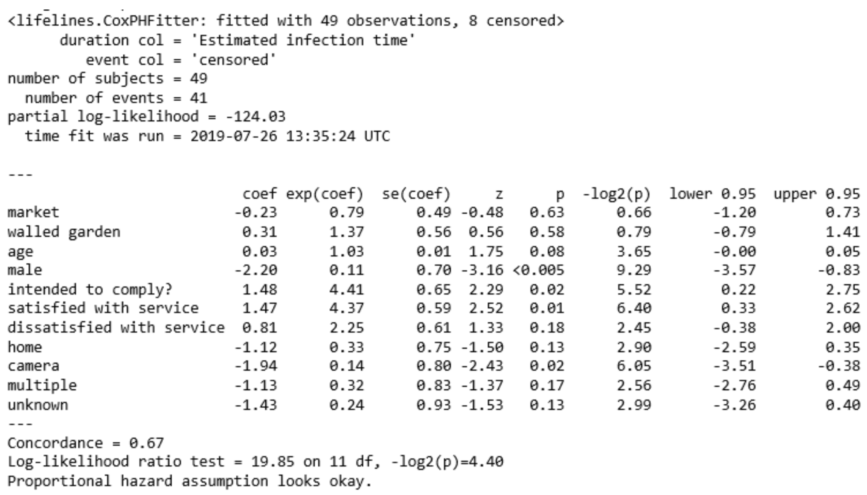
Model 1 is better than model 0

</td>
</tr>
<tr>
<td>Step 2</td>
<td>

The ‘right measures’ variable is the least reliable and excluded

Modeling the Cox model for [market, walled garden, age, male, intended to comply, satisfied with service, dissatisfied with service, home, camera, multiple, unknown]

</td>
</tr>
</table>



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -124.03
  time fit was run = 2019-07-26 13:35:24 UTC

---
                          coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
market                   -0.23      0.79      0.49 -0.48   0.63      0.66       -1.20        0.73
walled garden             0.31      1.37      0.56  0.56   0.58      0.79       -0.79        1.41
age                       0.03      1.03      0.01  1.75   0.08      3.65       -0.00        0.05
male                     -2.20      0.11      0.70 -3.16 <0.005      9.29       -3.57       -0.83
intended to comply?       1.48      4.41      0.65  2.29   0.02      5.52        0.22        2.75
satisfied with service    1.47      4.37      0.59  2.52   0.01      6.40        0.33        2.62
dissatisfied with service 0.81      2.25      0.61  1.33   0.18      2.45       -0.38        2.00
home                     -1.12      0.33      0.75 -1.50   0.13      2.90       -2.59        0.35
camera                   -1.94      0.14      0.80 -2.43   0.02      6.05       -3.51       -0.38
multiple                 -1.13      0.32      0.83 -1.37   0.17      2.56       -2.76        0.49
unknown                  -1.43      0.24      0.93 -1.53   0.13      2.99       -3.26        0.40
---
Concordance = 0.67
Log-likelihood ratio test = 19.85 on 11 df, -log2(p)=4.40
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,05

Log-Likelihood Ratio Statistic (LRS) = 0,1

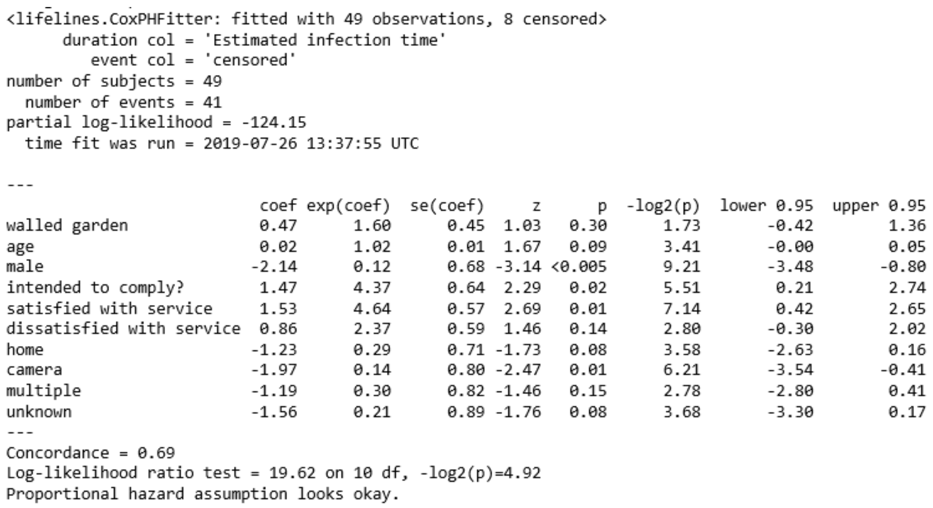
Consulting the chi-square distribution for 0,1 on 1 df: p>0,75

Model 2 is better than model 1

Step 3

The ‘market’ variable is the least reliable and excluded

Modeling the Cox model for [walled garden, age, male, intended to comply, satisfied with service, dissatisfied with service, home, camera, multiple, unknown]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -124.15
  time fit was run = 2019-07-26 13:37:55 UTC

---
                           coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
walled garden              0.47      1.60      0.45  1.03   0.30      1.73       -0.42        1.36
age                        0.02      1.02      0.01  1.67   0.09      3.41       -0.00        0.05
male                      -2.14      0.12      0.68 -3.14 <0.005      9.21       -3.48       -0.80
intended to comply?        1.47      4.37      0.64  2.29   0.02      5.51        0.21        2.74
satisfied with service     1.53      4.64      0.57  2.69   0.01      7.14        0.42        2.65
dissatisfied with service  0.86      2.37      0.59  1.46   0.14      2.80       -0.30        2.02
home                      -1.23      0.29      0.71 -1.73   0.08      3.58       -2.63        0.16
camera                    -1.97      0.14      0.80 -2.47   0.01      6.21       -3.54       -0.41
multiple                  -1.19      0.30      0.82 -1.46   0.15      2.78       -2.80        0.41
unknown                   -1.56      0.21      0.89 -1.76   0.08      3.68       -3.30        0.17
---
Concordance = 0.69
Log-likelihood ratio test = 19.62 on 10 df, -log2(p)=4.92
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,12

Log-Likelihood Ratio Statistic (LRS) = 0,24

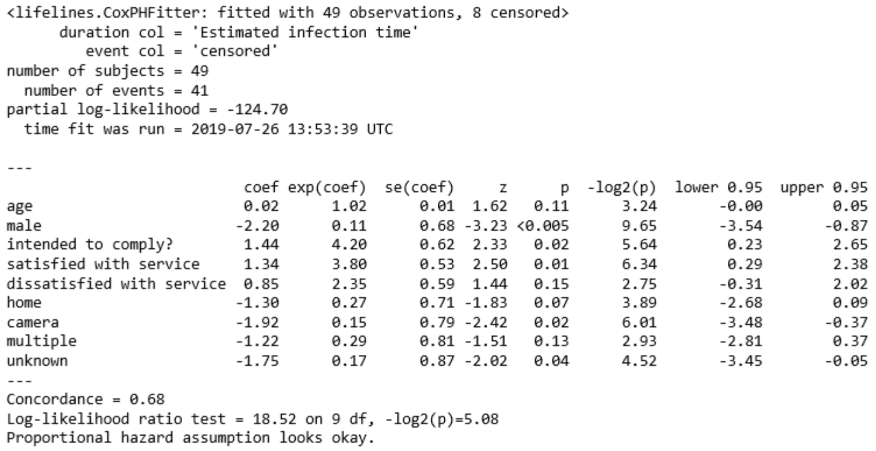
Consulting the chi-square distribution for 0,24 on 1 df: p>0,50

Model 3 is better than model 2

Step 4

The ‘walled garden’ variable is the least reliable and excluded

Modeling the Cox model for [age, male, intended to comply, satisfied with service, dissatisfied with service, home, camera, multiple, unknown]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -124.70
  time fit was run = 2019-07-26 13:53:39 UTC

---
                          coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
age                       0.02      1.02      0.01  1.62   0.11      3.24       -0.00        0.05
male                     -2.20      0.11      0.68 -3.23 <0.005      9.65       -3.54       -0.87
intended to comply?       1.44      4.20      0.62  2.33   0.02      5.64        0.23        2.65
satisfied with service    1.34      3.80      0.53  2.50   0.01      6.34        0.29        2.38
dissatisfied with service 0.85      2.35      0.59  1.44   0.15      2.75       -0.31        2.02
home                     -1.30      0.27      0.71 -1.83   0.07      3.89       -2.68        0.09
camera                   -1.92      0.15      0.79 -2.42   0.02      6.01       -3.48       -0.37
multiple                 -1.22      0.29      0.81 -1.51   0.13      2.93       -2.81        0.37
unknown                  -1.75      0.17      0.87 -2.02   0.04      4.52       -3.45       -0.05
---
Concordance = 0.68
Log-likelihood ratio test = 18.52 on 9 df, -log2(p)=5.08
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,55

Log-Likelihood Ratio Statistic (LRS) = 1,1

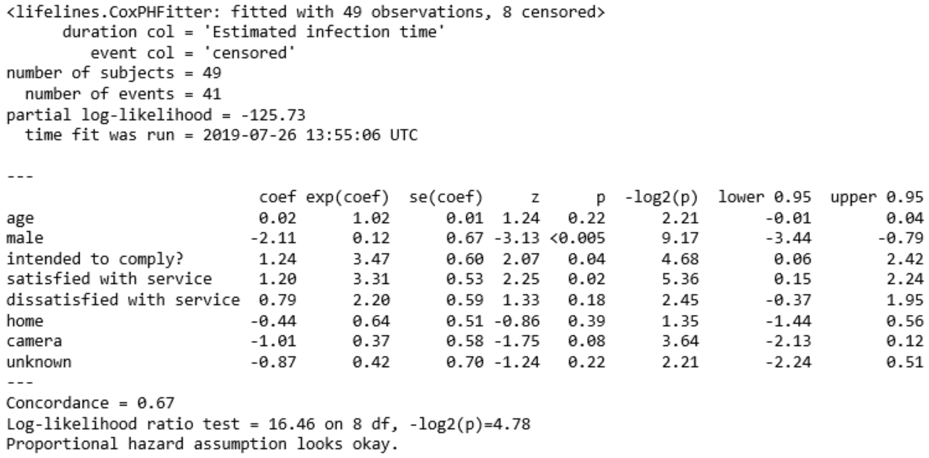
Consulting the chi-square distribution for 1,1 on 1 df: p>0,25

Model 4 is better than model 3

Step 5

The ‘multiple’ variable is the least reliable and excluded

Modeling the Cox model for [age, male, intended to comply, satisfied with service, dissatisfied with service, home, camera, unknown]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -125.73
  time fit was run = 2019-07-26 13:55:06 UTC

---
                          coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
age                       0.02      1.02      0.01  1.24   0.22      2.21       -0.01        0.04
male                     -2.11      0.12      0.67 -3.13 <0.005      9.17       -3.44       -0.79
intended to comply?       1.24      3.47      0.60  2.07   0.04      4.68        0.06        2.42
satisfied with service    1.20      3.31      0.53  2.25   0.02      5.36        0.15        2.24
dissatisfied with service 0.79      2.20      0.59  1.33   0.18      2.45       -0.37        1.95
home                     -0.44      0.64      0.51 -0.86   0.39      1.35       -1.44        0.56
camera                   -1.01      0.37      0.58 -1.75   0.08      3.64       -2.13        0.12
unknown                  -0.87      0.42      0.70 -1.24   0.22      2.21       -2.24        0.51
---
Concordance = 0.67
Log-likelihood ratio test = 16.46 on 8 df, -log2(p)=4.78
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -1,03

Log-Likelihood Ratio Statistic (LRS) = 2,06

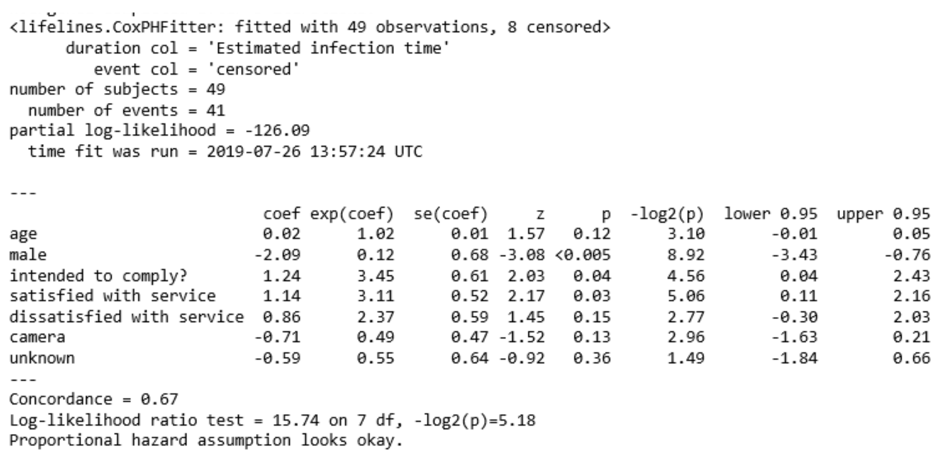
Consulting the chi-square distribution for 2,06 on 1 df: p>0,10

Model 5 is better than model 4

Step 6

The ‘home’ variable is the least reliable and excluded

Modeling the Cox model for [age, male, intended to comply, satisfied with service, dissatisfied with service, camera, unknown]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -126.09
  time fit was run = 2019-07-26 13:57:24 UTC

---
                           coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
age                        0.02      1.02      0.01  1.57   0.12      3.10       -0.01        0.05
male                      -2.09      0.12      0.68 -3.08 <0.005      8.92       -3.43       -0.76
intended to comply?        1.24      3.45      0.61  2.03   0.04      4.56        0.04        2.43
satisfied with service     1.14      3.11      0.52  2.17   0.03      5.06        0.11        2.16
dissatisfied with service  0.86      2.37      0.59  1.45   0.15      2.77       -0.30        2.03
camera                    -0.71      0.49      0.47 -1.52   0.13      2.96       -1.63        0.21
unknown                   -0.59      0.55      0.64 -0.92   0.36      1.49       -1.84        0.66
---
Concordance = 0.67
Log-likelihood ratio test = 15.74 on 7 df, -log2(p)=5.18
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,36

Log-Likelihood Ratio Statistic (LRS) = 0,72

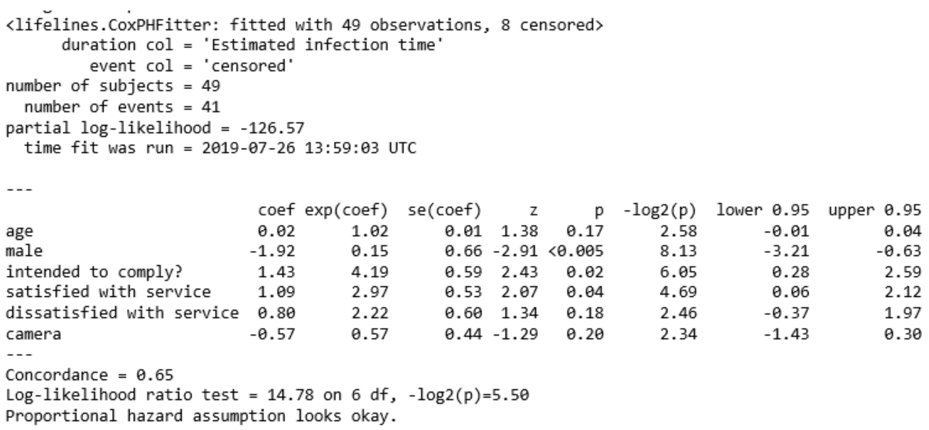
Consulting the chi-square distribution for 0,72 on 1 df: p>0,25

Model 6 is better than model 5

Step 7

The ‘unknown’ variable is the least reliable and excluded

Modeling the Cox model for [age, male, intended to comply, satisfied with service, dissatisfied with service, camera]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -126.57
  time fit was run = 2019-07-26 13:59:03 UTC

---
                           coef exp(coef)  se(coef)     z      p  -log2(p)  lower 0.95  upper 0.95
age                        0.02      1.02      0.01  1.38   0.17      2.58       -0.01        0.04
male                      -1.92      0.15      0.66 -2.91 <0.005      8.13       -3.21       -0.63
intended to comply?        1.43      4.19      0.59  2.43   0.02      6.05        0.28        2.59
satisfied with service     1.09      2.97      0.53  2.07   0.04      4.69        0.06        2.12
dissatisfied with service  0.80      2.22      0.60  1.34   0.18      2.46       -0.37        1.97
camera                    -0.57      0.57      0.44 -1.29   0.20      2.34       -1.43        0.30
---
Concordance = 0.65
Log-likelihood ratio test = 14.78 on 6 df, -log2(p)=5.50
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,48

Log-Likelihood Ratio Statistic (LRS) = 0,96

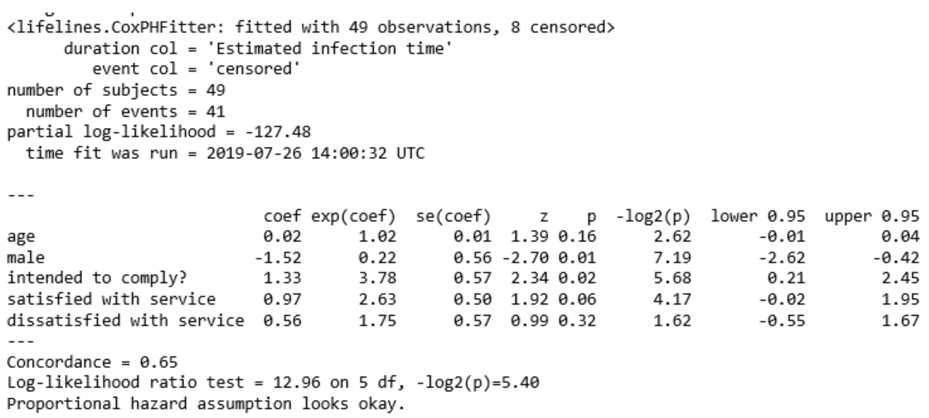
Consulting the chi-square distribution for 0,96 on 1 df: p>0,25

Model 7 is better than model 6

Step 8

The ‘camera’ variable is the least reliable and excluded

Modeling the Cox model for [age, male, intended to comply, satisfied with service, dissatisfied with service]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -127.48
  time fit was run = 2019-07-26 14:00:32 UTC

---
                          coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
age                       0.02      1.02      0.01  1.39 0.16      2.62       -0.01        0.04
male                     -1.52      0.22      0.56 -2.70 0.01      7.19       -2.62       -0.42
intended to comply?       1.33      3.78      0.57  2.34 0.02      5.68        0.21        2.45
satisfied with service    0.97      2.63      0.50  1.92 0.06      4.17       -0.02        1.95
dissatisfied with service 0.56      1.75      0.57  0.99 0.32      1.62       -0.55        1.67
---
Concordance = 0.65
Log-likelihood ratio test = 12.96 on 5 df, -log2(p)=5.40
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,91

Log-Likelihood Ratio Statistic (LRS) = 1,82

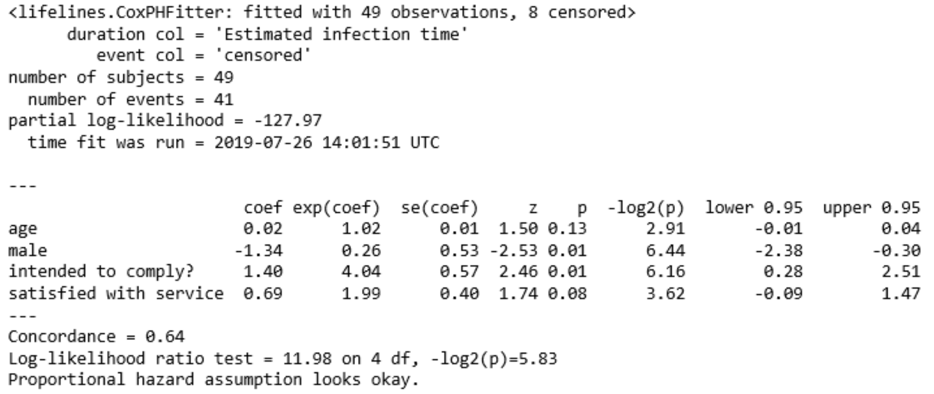
Consulting the chi-square distribution for 1,82 on 1 df: p>0,10

Model 8 is better than model 7

Step 9

The ‘dissatisfied with service’ variable is the least reliable and excluded

Modeling the Cox model for [age, male, intended to comply, satisfied with service]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -127.97
  time fit was run = 2019-07-26 14:01:51 UTC

---
                       coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
age                    0.02      1.02      0.01  1.50 0.13      2.91       -0.01        0.04
male                  -1.34      0.26      0.53 -2.53 0.01      6.44       -2.38       -0.30
intended to comply?    1.40      4.04      0.57  2.46 0.01      6.16        0.28        2.51
satisfied with service 0.69      1.99      0.40  1.74 0.08      3.62       -0.09        1.47
---
Concordance = 0.64
Log-likelihood ratio test = 11.98 on 4 df, -log2(p)=5.83
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,49

Log-Likelihood Ratio Statistic (LRS) = 0,98

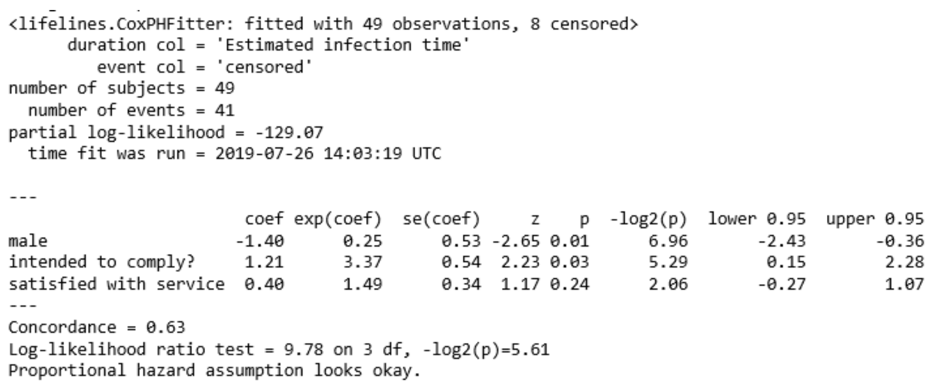
Consulting the chi-square distribution for 0,98 on 1 df: p>0,25

Model 9 is better than model 8

Step 10

The ‘age’ variable is the least reliable and excluded

Modeling the Cox model for [male, intended to comply, satisfied with service]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -129.07
  time fit was run = 2019-07-26 14:03:19 UTC

---
                       coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
male                  -1.40      0.25      0.53 -2.65 0.01      6.96       -2.43       -0.36
intended to comply?    1.21      3.37      0.54  2.23 0.03      5.29        0.15        2.28
satisfied with service 0.40      1.49      0.34  1.17 0.24      2.06       -0.27        1.07
---
Concordance = 0.63
Log-likelihood ratio test = 9.78 on 3 df, -log2(p)=5.61
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -1,1

Log-Likelihood Ratio Statistic (LRS) = 2,2

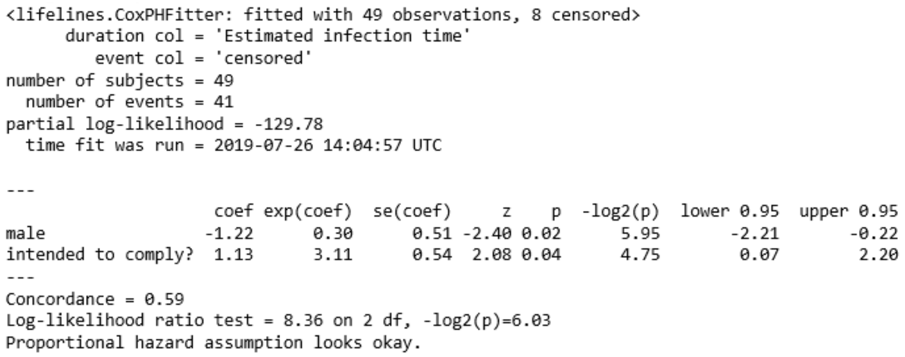
Consulting the chi-square distribution for 2,2 on 1 df: p>0,10

Model 10 is better than model 9

Step 11

The ‘satisfied with service’ variable is the least reliable and excluded

Modeling the Cox model for [male, intended to comply]



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -129.78
  time fit was run = 2019-07-26 14:04:57 UTC

---
                     coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
male                -1.22      0.30      0.51 -2.40 0.02      5.95       -2.21       -0.22
intended to comply?  1.13      3.11      0.54  2.08 0.04      4.75        0.07        2.20
---
Concordance = 0.59
Log-likelihood ratio test = 8.36 on 2 df, -log2(p)=6.03
Proportional hazard assumption looks okay.
```

The difference in partial log-likelihood (LL) = -0,71

Log-Likelihood Ratio Statistic (LRS) = 1,42

Consulting the chi-square distribution for 1,42 on 1 df: p>0,10

Model 11 is better than model 10

Step 12

All variables are significant. Exclusion of ‘intended to comply’ leads to a Log-Likelihood of -132,61

The difference in partial log-likelihood (LL) = -2,83

Log-Likelihood Ratio Statistic (LRS) = 5,66

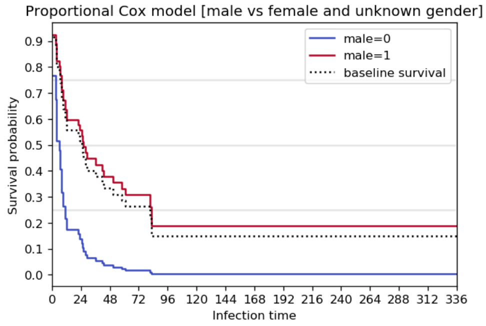
Consulting the chi-square distribution 5,66 on 1 df: p<0,05

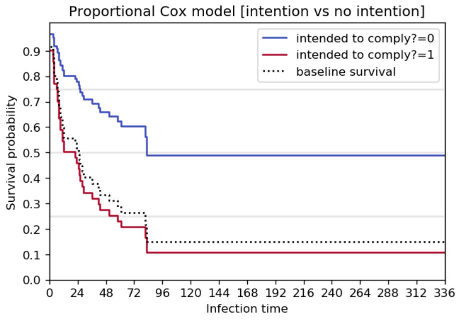
Model 11 is best of all models.

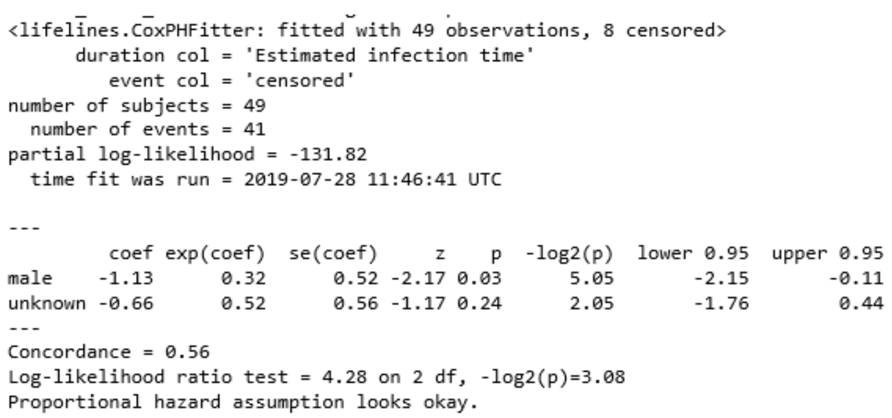
When comparing model 11 with a trivial model, the LRS is 8,36 for 2 degrees of freedom: p<0,05

Model 11 is better than a model without covariates.

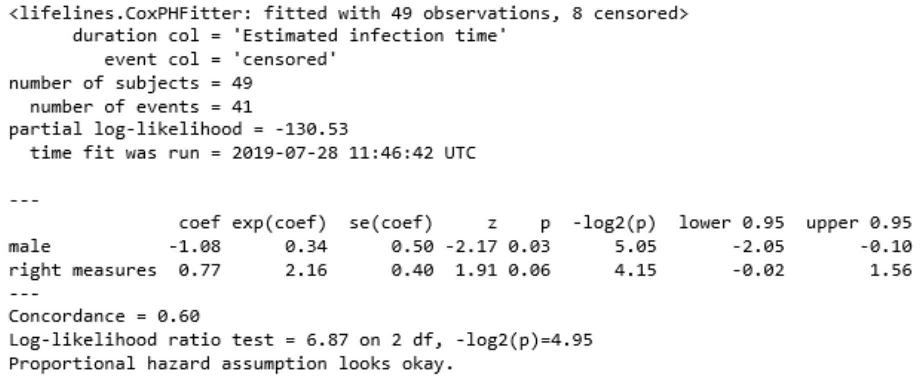
Model 11 is accepted



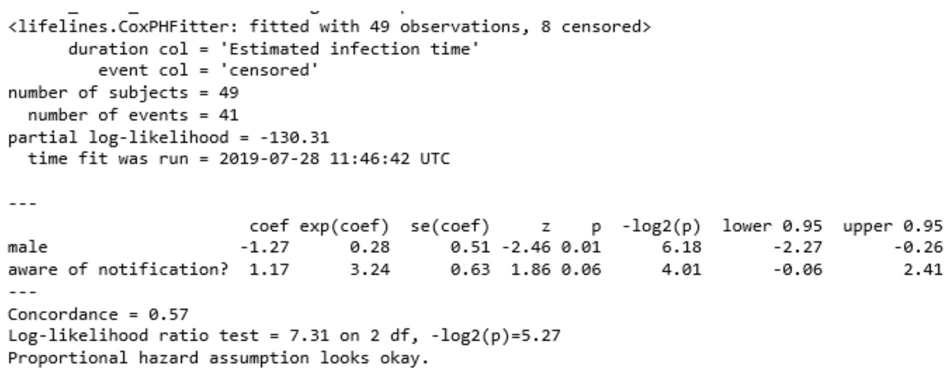




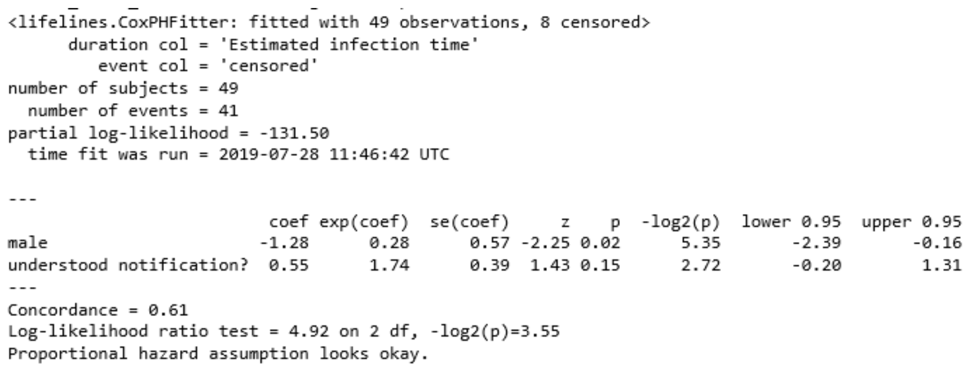
```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -131.82
  time fit was run = 2019-07-28 11:46:41 UTC

---
         coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
male    -1.13      0.32      0.52 -2.17 0.03      5.05       -2.15       -0.11
unknown -0.66      0.52      0.56 -1.17 0.24      2.05       -1.76        0.44
---
Concordance = 0.56
Log-likelihood ratio test = 4.28 on 2 df, -log2(p)=3.08
Proportional hazard assumption looks okay.
```



```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -130.53
  time fit was run = 2019-07-28 11:46:42 UTC

---
               coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
male          -1.08      0.34      0.50 -2.17 0.03      5.05       -2.05       -0.10
right measures  0.77      2.16      0.40  1.91 0.06      4.15       -0.02        1.56
---
Concordance = 0.60
Log-likelihood ratio test = 6.87 on 2 df, -log2(p)=4.95
Proportional hazard assumption looks okay.
```



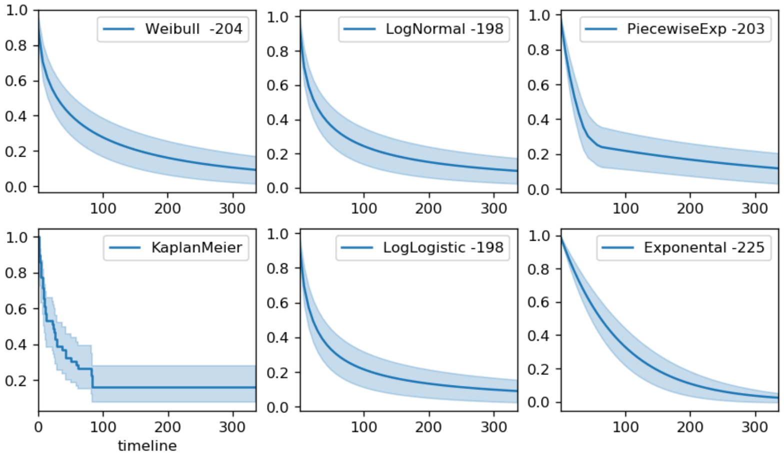
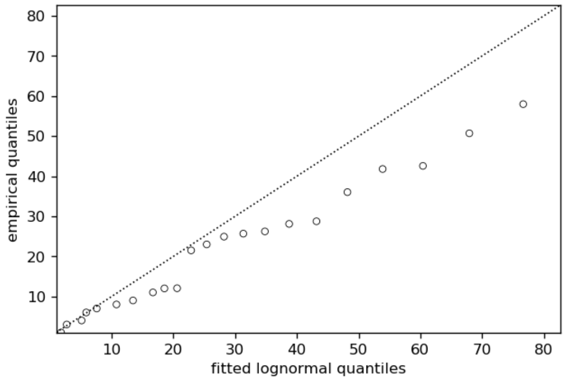
```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -130.31
  time fit was run = 2019-07-28 11:46:42 UTC

---
                       coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
male                  -1.27      0.28      0.51 -2.46 0.01      6.18       -2.27       -0.26
aware of notification?  1.17      3.24      0.63  1.86 0.06      4.01       -0.06        2.41
---
Concordance = 0.57
Log-likelihood ratio test = 7.31 on 2 df, -log2(p)=5.27
Proportional hazard assumption looks okay.
```

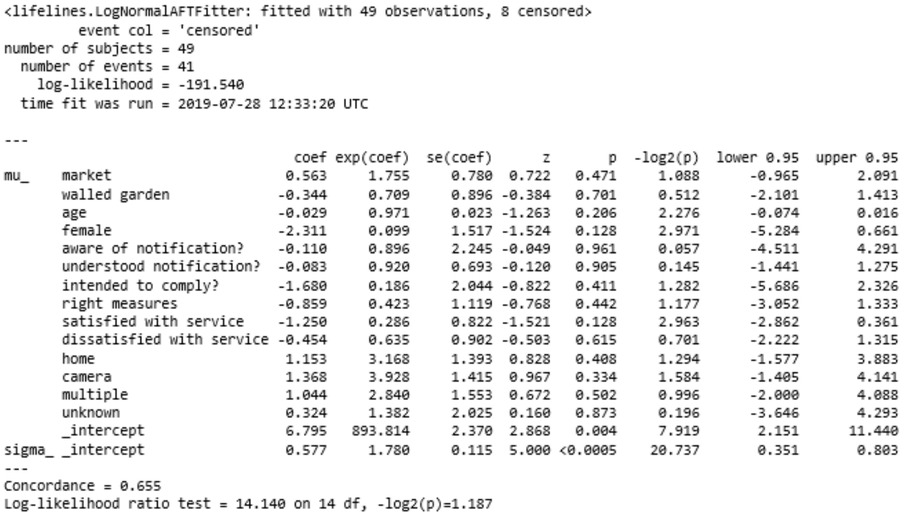


```
<lifelines.CoxPHFitter: fitted with 49 observations, 8 censored>
      duration col = 'Estimated infection time'
         event col = 'censored'
number of subjects = 49
  number of events = 41
partial log-likelihood = -131.50
  time fit was run = 2019-07-28 11:46:42 UTC

---
                          coef exp(coef)  se(coef)     z    p  -log2(p)  lower 0.95  upper 0.95
male                     -1.28      0.28      0.57 -2.25 0.02      5.35       -2.39       -0.16
understood notification?  0.55      1.74      0.39  1.43 0.15      2.72       -0.20        1.31
---
Concordance = 0.61
Log-likelihood ratio test = 4.92 on 2 df, -log2(p)=3.55
Proportional hazard assumption looks okay.
```

Model 11 has a Log-Likelihood of -129,78, which is higher than these models with substituted variables for the variable intention. Model 11 is thus the best model.

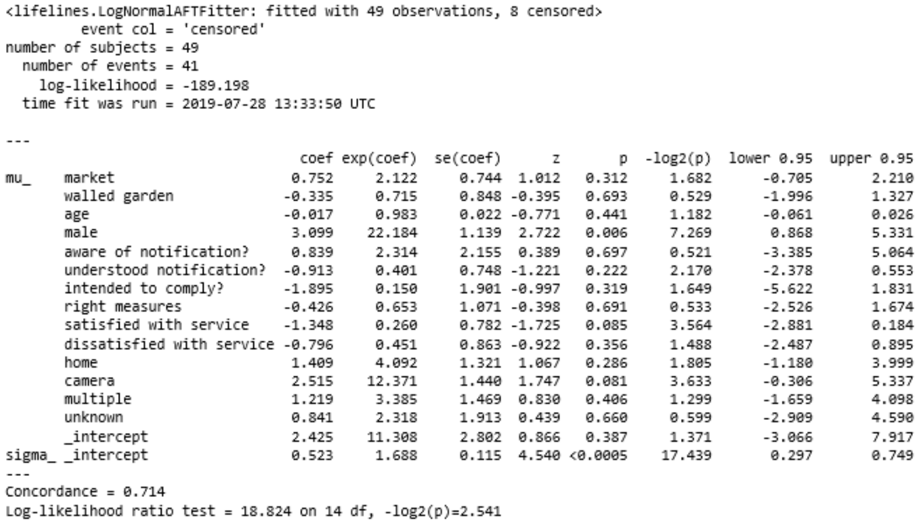
L.4 AFT modeling steps

Step 0	Following the similar line of reasoning of appendix J.4, the fitted distributions in figure 60 show that both the LogNormal and LogLogistic distributions have the best fit. The Lognormal distribution has a slightly little better fit (LL = 198,61) than the LogLogistic distribution (-198,63). We, therefore, use the LogNormal distribution for the AFT model.  Figure 58 Distribution fits for the survival curve of all observations Figure 61 shows the quantile-quantile (Q-Q) plot to compare the fitted LogNormal distribution with the empirical distribution. The empirical distribution more concentrated than the fitted distribution and thus has a larger tail than the fitted LogNormal distribution. This is also visible through the data range: almost all data points lie between zero and eighty hours.  Figure 59 Q-Q plot LogNormal distribution
Step 1	Similar to the Cox modeling in the previous section, we first estimate two models (including female and including male) so we can decide which variable to continue with. Including female:



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -191.540
  time fit was run = 2019-07-28 12:33:20 UTC

---
                                 coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                   0.563     1.755     0.780  0.722   0.471     1.088      -0.965       2.091
       walled garden           -0.344     0.709     0.896 -0.384   0.701     0.512      -2.101       1.413
       age                     -0.029     0.971     0.023 -1.263   0.206     2.276      -0.074       0.016
       female                  -2.311     0.099     1.517 -1.524   0.128     2.971      -5.284       0.661
       aware of notification?  -0.110     0.896     2.245 -0.049   0.961     0.057      -4.511       4.291
       understood notification? -0.083    0.920     0.693 -0.120   0.905     0.145      -1.441       1.275
       intended to comply?     -1.680     0.186     2.044 -0.822   0.411     1.282      -5.686       2.326
       right measures          -0.859     0.423     1.119 -0.768   0.442     1.177      -3.052       1.333
       satisfied with service  -1.250     0.286     0.822 -1.521   0.128     2.963      -2.862       0.361
       dissatisfied with service -0.454   0.635     0.902 -0.503   0.615     0.701      -2.222       1.315
       home                     1.153     3.168     1.393  0.828   0.408     1.294      -1.577       3.883
       camera                   1.368     3.928     1.415  0.967   0.334     1.584      -1.405       4.141
       multiple                 1.044     2.840     1.553  0.672   0.502     0.996      -2.000       4.088
       unknown                  0.324     1.382     2.025  0.160   0.873     0.196      -3.646       4.293
       _intercept               6.795   893.814     2.370  2.868   0.004     7.919       2.151      11.440
sigma_ _intercept               0.577     1.780     0.115  5.000 <0.0005    20.737       0.351       0.803
---
Concordance = 0.655
Log-likelihood ratio test = 14.140 on 14 df, -log2(p)=1.187
```

Including male:



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -189.198
  time fit was run = 2019-07-28 13:33:50 UTC

---
                                 coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                   0.752     2.122     0.744  1.012   0.312     1.682      -0.705       2.210
       walled garden           -0.335     0.715     0.848 -0.395   0.693     0.529      -1.996       1.327
       age                     -0.017     0.983     0.022 -0.771   0.441     1.182      -0.061       0.026
       male                     3.099    22.184     1.139  2.722   0.006     7.269       0.868       5.331
       aware of notification?   0.839     2.314     2.155  0.389   0.697     0.521      -3.385       5.064
       understood notification? -0.913    0.401     0.748 -1.221   0.222     2.170      -2.378       0.553
       intended to comply?     -1.895     0.150     1.901 -0.997   0.319     1.649      -5.622       1.831
       right measures          -0.426     0.653     1.071 -0.398   0.691     0.533      -2.526       1.674
       satisfied with service  -1.348     0.260     0.782 -1.725   0.085     3.564      -2.881       0.184
       dissatisfied with service -0.796   0.451     0.863 -0.922   0.356     1.488      -2.487       0.895
       home                     1.409     4.092     1.321  1.067   0.286     1.805      -1.180       3.999
       camera                   2.515    12.371     1.440  1.747   0.081     3.633      -0.306       5.337
       multiple                 1.219     3.385     1.469  0.830   0.406     1.299      -1.659       4.098
       unknown                  0.841     2.318     1.913  0.439   0.660     0.599      -2.909       4.590
       _intercept               2.425    11.308     2.802  0.866   0.387     1.371      -3.066       7.917
sigma_ _intercept               0.523     1.688     0.115  4.540 <0.0005    17.439       0.297       0.749
---
Concordance = 0.714
Log-likelihood ratio test = 18.824 on 14 df, -log2(p)=2.541
```

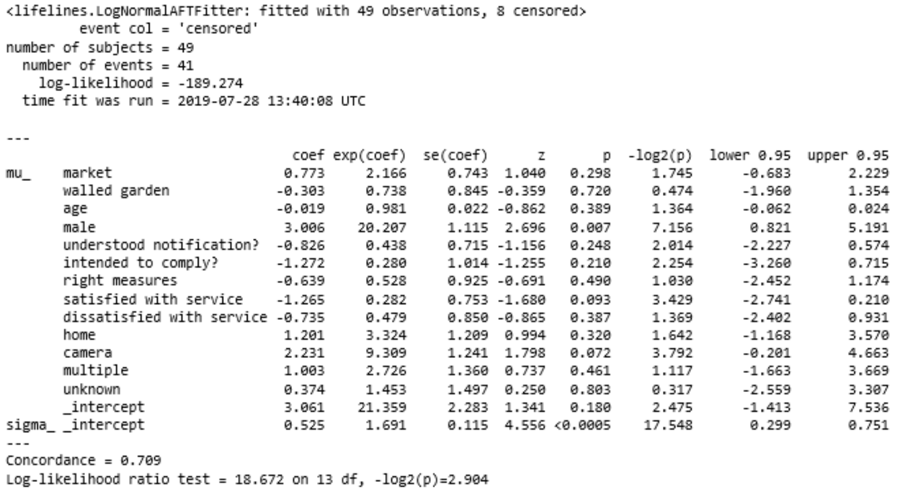
The model including ‘male’ has higher reliable parameters and a better overall fit (the partial log-likelihood is higher for equal degrees of freedom).

In the next steps, the variable ‘female’ is excluded. Due to this exclusion, the coding of the dummy variable ‘male’ is changed. 1 = male subscriber, and 0 = female subscriber and subscribers of unknown gender.

Step 2

The ‘aware of notification’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, walled garden, age, male, understood notification, intended to comply, right measures, sastisfied with service, dissastisfied with service, home, camera, multiple, unknown]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
        event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -189.274
  time fit was run = 2019-07-28 13:40:08 UTC

---
                               coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                 0.773     2.166     0.743  1.040   0.298     1.745      -0.683       2.229
       walled garden         -0.303     0.738     0.845 -0.359   0.720     0.474      -1.960       1.354
       age                   -0.019     0.981     0.022 -0.862   0.389     1.364      -0.062       0.024
       male                   3.006    20.207     1.115  2.696   0.007     7.156       0.821       5.191
       understood notification? -0.826  0.438     0.715 -1.156   0.248     2.014      -2.227       0.574
       intended to comply?   -1.272     0.280     1.014 -1.255   0.210     2.254      -3.260       0.715
       right measures        -0.639     0.528     0.925 -0.691   0.490     1.030      -2.452       1.174
       satisfied with service -1.265    0.282     0.753 -1.680   0.093     3.429      -2.741       0.210
       dissatisfied with service -0.735 0.479     0.850 -0.865   0.387     1.369      -2.402       0.931
       home                   1.201     3.324     1.209  0.994   0.320     1.642      -1.168       3.570
       camera                 2.231     9.309     1.241  1.798   0.072     3.792      -0.201       4.663
       multiple               1.003     2.726     1.360  0.737   0.461     1.117      -1.663       3.669
       unknown                0.374     1.453     1.497  0.250   0.803     0.317      -2.559       3.307
       _intercept             3.061    21.359     2.283  1.341   0.180     2.475      -1.413       7.536
sigma_ _intercept             0.525     1.691     0.115  4.556 <0.0005    17.548       0.299       0.751
---
Concordance = 0.709
Log-likelihood ratio test = 18.672 on 13 df, -log2(p)=2.904
```

The difference in partial log-likelihood (LL) = -0,076

Log-Likelihood Ratio Statistic (LRS) = 0,152

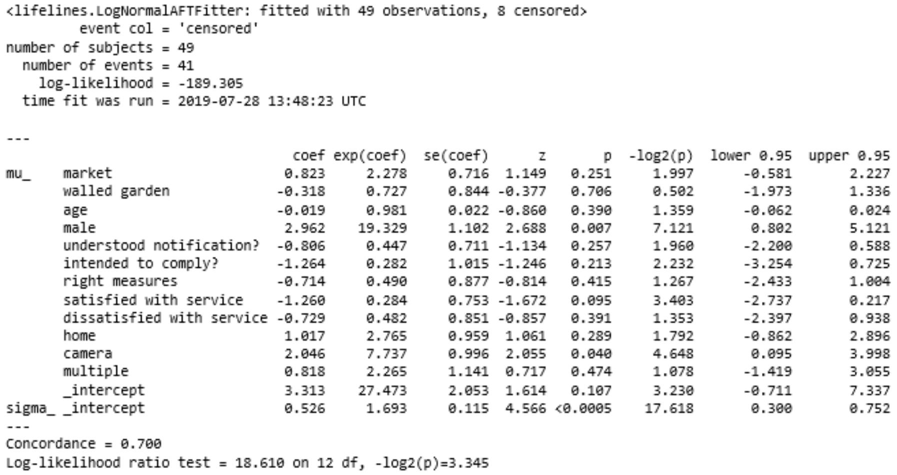
Consulting the chi-square distribution for 0,152 on 1 df: p>0,50

Model 2 is better than model 1

Step3

The ‘unknown’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, walled garden, age, male, understood notification, intended to comply, right measures, sastisfied with service, dissastisfied with service, home, camera, multiple]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
        event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -189.305
  time fit was run = 2019-07-28 13:48:23 UTC

---
                               coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                 0.823     2.278     0.716  1.149   0.251     1.997      -0.581       2.227
       walled garden         -0.318     0.727     0.844 -0.377   0.706     0.502      -1.973       1.336
       age                   -0.019     0.981     0.022 -0.860   0.390     1.359      -0.062       0.024
       male                   2.962    19.329     1.102  2.688   0.007     7.121       0.802       5.121
       understood notification? -0.806  0.447     0.711 -1.134   0.257     1.960      -2.200       0.588
       intended to comply?   -1.264     0.282     1.015 -1.246   0.213     2.232      -3.254       0.725
       right measures        -0.714     0.490     0.877 -0.814   0.415     1.267      -2.433       1.004
       satisfied with service -1.260    0.284     0.753 -1.672   0.095     3.403      -2.737       0.217
       dissatisfied with service -0.729 0.482     0.851 -0.857   0.391     1.353      -2.397       0.938
       home                   1.017     2.765     0.959  1.061   0.289     1.792      -0.862       2.896
       camera                 2.046     7.737     0.996  2.055   0.040     4.648       0.095       3.998
       multiple               0.818     2.265     1.141  0.717   0.474     1.078      -1.419       3.055
       _intercept             3.313    27.473     2.053  1.614   0.107     3.230      -0.711       7.337
sigma_ _intercept             0.526     1.693     0.115  4.566 <0.0005    17.618       0.300       0.752
---
Concordance = 0.700
Log-likelihood ratio test = 18.610 on 12 df, -log2(p)=3.345
```

The difference in partial log-likelihood (LL) = -0,031

Log-Likelihood Ratio Statistic (LRS) = 0,062

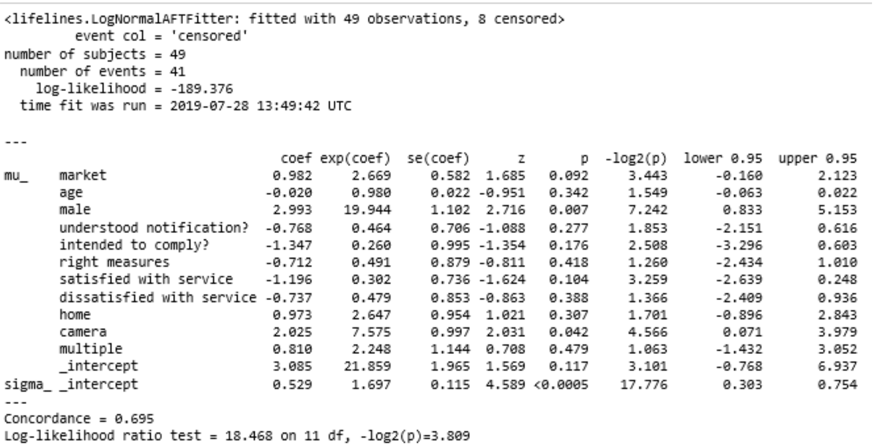
Consulting the chi-square distribution for 0,062 on 1 df: p>0,75

Model 3 is better than model 2

Step 4

The ‘walled garden’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, age, male, understood notification, intended to comply, right measures, sastisfied with service, dissastisfied with service, home, camera, multiple]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
        event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -189.376
  time fit was run = 2019-07-28 13:49:42 UTC

---
                                 coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                   0.982     2.669     0.582  1.685   0.092     3.443      -0.160       2.123
       age                     -0.020     0.980     0.022 -0.951   0.342     1.549      -0.063       0.022
       male                     2.993    19.944     1.102  2.716   0.007     7.242       0.833       5.153
       understood notification? -0.768     0.464     0.706 -1.088   0.277     1.853      -2.151       0.616
       intended to comply?     -1.347     0.260     0.995 -1.354   0.176     2.508      -3.296       0.603
       right measures          -0.712     0.491     0.879 -0.811   0.418     1.260      -2.434       1.010
       satisfied with service  -1.196     0.302     0.736 -1.624   0.104     3.259      -2.639       0.248
       dissatisfied with service -0.737     0.479     0.853 -0.863   0.388     1.366      -2.409       0.936
       home                     0.973     2.647     0.954  1.021   0.307     1.701      -0.896       2.843
       camera                   2.025     7.575     0.997  2.031   0.042     4.566       0.071       3.979
       multiple                 0.810     2.248     1.144  0.708   0.479     1.063      -1.432       3.052
       _intercept               3.085    21.859     1.965  1.569   0.117     3.101      -0.768       6.937
sigma_ _intercept               0.529     1.697     0.115  4.589 <0.0005    17.776       0.303       0.754
---
Concordance = 0.695
Log-likelihood ratio test = 18.468 on 11 df, -log2(p)=3.809
```

The difference in partial log-likelihood (LL) = -0,071

Log-Likelihood Ratio Statistic (LRS) = 0,142

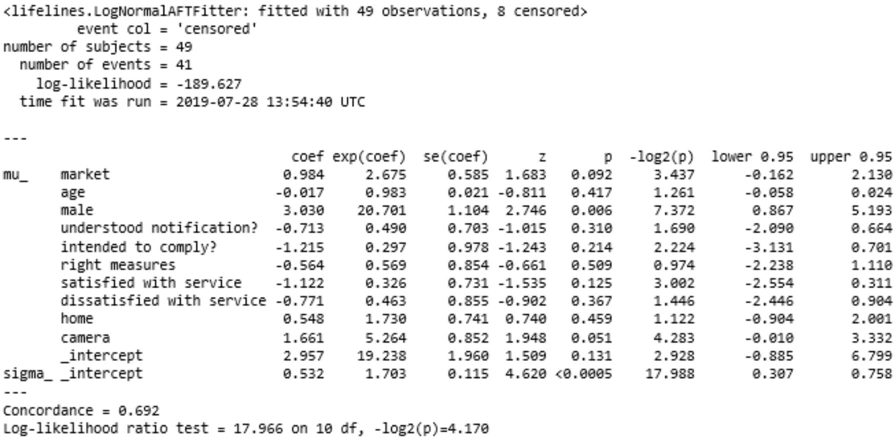
Consulting the chi-square distribution for 0,142 on 1 df: p>0,50

Model 4 is better than model 3

Step 5

The ‘multiple’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, age, male, understood notification, intended to comply, right measures, sastisfied with service, dissastisfied with service, home, camera]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
        event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -189.627
  time fit was run = 2019-07-28 13:54:40 UTC

---
                                 coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                   0.984     2.675     0.585  1.683   0.092     3.437      -0.162       2.130
       age                     -0.017     0.983     0.021 -0.811   0.417     1.261      -0.058       0.024
       male                     3.030    20.701     1.104  2.746   0.006     7.372       0.867       5.193
       understood notification? -0.713     0.490     0.703 -1.015   0.310     1.690      -2.090       0.664
       intended to comply?     -1.215     0.297     0.978 -1.243   0.214     2.224      -3.131       0.701
       right measures          -0.564     0.569     0.854 -0.661   0.509     0.974      -2.238       1.110
       satisfied with service  -1.122     0.326     0.731 -1.535   0.125     3.002      -2.554       0.311
       dissatisfied with service -0.771     0.463     0.855 -0.902   0.367     1.446      -2.446       0.904
       home                     0.548     1.730     0.741  0.740   0.459     1.122      -0.904       2.001
       camera                   1.661     5.264     0.852  1.948   0.051     4.283      -0.010       3.332
       _intercept               2.957    19.238     1.960  1.509   0.131     2.928      -0.885       6.799
sigma_ _intercept               0.532     1.703     0.115  4.620 <0.0005    17.988       0.307       0.758
---
Concordance = 0.692
Log-likelihood ratio test = 17.966 on 10 df, -log2(p)=4.170
```

The difference in partial log-likelihood (LL) = -0,251

Log-Likelihood Ratio Statistic (LRS) = 0,502

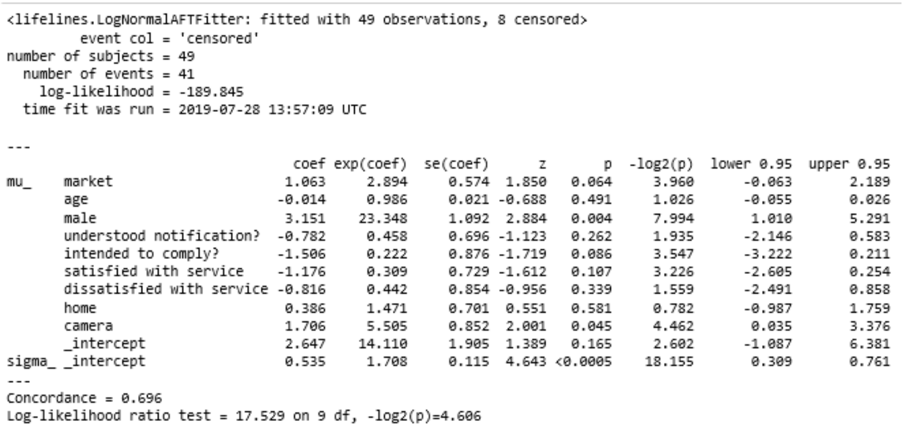
Consulting the chi-square distribution for 0,502 on 1 df: p>0,25

Model 5 is better than model 4

Step 6

The ‘right measure’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, age, male, understood notification, intended to comply, sastisfied with service, dissastisfied with service, home, camera]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -189.845
  time fit was run = 2019-07-28 13:57:09 UTC

---
                                  coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                    1.063     2.894     0.574  1.850   0.064     3.960      -0.063       2.189
       age                      -0.014     0.986     0.021 -0.688   0.491     1.026      -0.055       0.026
       male                      3.151    23.348     1.092  2.884   0.004     7.994       1.010       5.291
       understood notification? -0.782     0.458     0.696 -1.123   0.262     1.935      -2.146       0.583
       intended to comply?      -1.506     0.222     0.876 -1.719   0.086     3.547      -3.222       0.211
       satisfied with service   -1.176     0.309     0.729 -1.612   0.107     3.226      -2.605       0.254
       dissatisfied with service -0.816    0.442     0.854 -0.956   0.339     1.559      -2.491       0.858
       home                      0.386     1.471     0.701  0.551   0.581     0.782      -0.987       1.759
       camera                    1.706     5.505     0.852  2.001   0.045     4.462       0.035       3.376
       _intercept                2.647    14.110     1.905  1.389   0.165     2.602      -1.087       6.381
sigma_ _intercept                0.535     1.708     0.115  4.643 <0.0005    18.155       0.309       0.761
---
Concordance = 0.696
Log-likelihood ratio test = 17.529 on 9 df, -log2(p)=4.606
```

The difference in partial log-likelihood (LL) = -0,218

Log-Likelihood Ratio Statistic (LRS) = 0,436

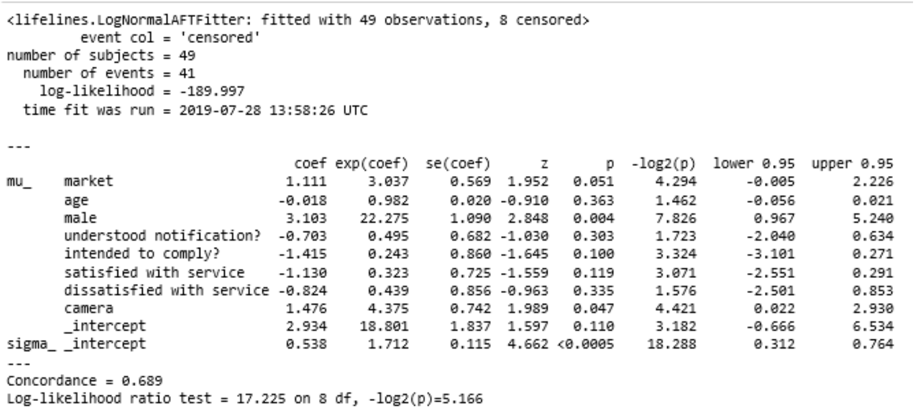
Consulting the chi-square distribution for 0,436 on 1 df: p>0,50

Model 6 is better than model 5

Step 7

The ‘home’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, age, male, understood notification, intended to comply, sastisfied with service, dissastisfied with service, camera]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -189.997
  time fit was run = 2019-07-28 13:58:26 UTC

---
                                  coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                    1.111     3.037     0.569  1.952   0.051     4.294      -0.005       2.226
       age                      -0.018     0.982     0.020 -0.910   0.363     1.462      -0.056       0.021
       male                      3.103    22.275     1.090  2.848   0.004     7.826       0.967       5.240
       understood notification? -0.703     0.495     0.682 -1.030   0.303     1.723      -2.040       0.634
       intended to comply?      -1.415     0.243     0.860 -1.645   0.100     3.324      -3.101       0.271
       satisfied with service   -1.130     0.323     0.725 -1.559   0.119     3.071      -2.551       0.291
       dissatisfied with service -0.824    0.439     0.856 -0.963   0.335     1.576      -2.501       0.853
       camera                    1.476     4.375     0.742  1.989   0.047     4.421       0.022       2.930
       _intercept                2.934    18.801     1.837  1.597   0.110     3.182      -0.666       6.534
sigma_ _intercept                0.538     1.712     0.115  4.662 <0.0005    18.288       0.312       0.764
---
Concordance = 0.689
Log-likelihood ratio test = 17.225 on 8 df, -log2(p)=5.166
```

The difference in partial log-likelihood (LL) = -0,152

Log-Likelihood Ratio Statistic (LRS) = 0,304

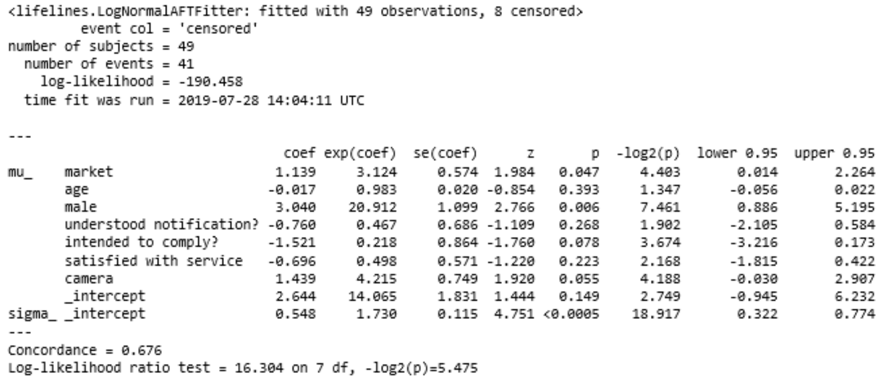
Consulting the chi-square distribution for 0,304 on 1 df: p>0,50

Model 7 is better than model 6

Step 8

The ‘dissatisfied with service’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, age, male, understood notification, intended to comply, sastisfied with service, camera]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -190.458
  time fit was run = 2019-07-28 14:04:11 UTC

---
                                coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                  1.139     3.124     0.574  1.984   0.047     4.403       0.014       2.264
       age                    -0.017     0.983     0.020 -0.854   0.393     1.347      -0.056       0.022
       male                    3.040    20.912     1.099  2.766   0.006     7.461       0.886       5.195
       understood notification? -0.760   0.467     0.686 -1.109   0.268     1.902      -2.105       0.584
       intended to comply?    -1.521     0.218     0.864 -1.760   0.078     3.674      -3.216       0.173
       satisfied with service -0.696     0.498     0.571 -1.220   0.223     2.168      -1.815       0.422
       camera                  1.439     4.215     0.749  1.920   0.055     4.188      -0.030       2.907
       _intercept              2.644    14.065     1.831  1.444   0.149     2.749      -0.945       6.232
sigma_ _intercept              0.548     1.730     0.115  4.751 <0.0005    18.917       0.322       0.774
---
Concordance = 0.676
Log-likelihood ratio test = 16.304 on 7 df, -log2(p)=5.475
```

The difference in partial log-likelihood (LL) = -0,461

Log-Likelihood Ratio Statistic (LRS) = 0,922

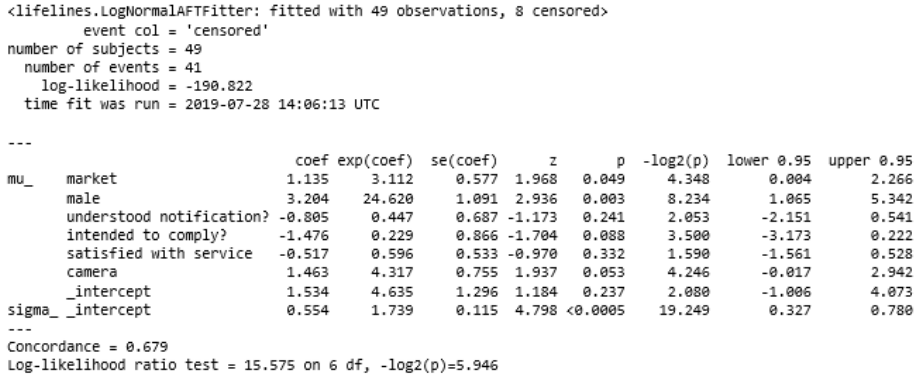
Consulting the chi-square distribution for 0,922 on 1 df: p>0,25

Model 8 is better than model 7

Step 9

The ‘age’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, male, understood notification, intended to comply, sastisfied with service, camera]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -190.822
  time fit was run = 2019-07-28 14:06:13 UTC

---
                                coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                  1.135     3.112     0.577  1.968   0.049     4.348       0.004       2.266
       male                    3.204    24.620     1.091  2.936   0.003     8.234       1.065       5.342
       understood notification? -0.805   0.447     0.687 -1.173   0.241     2.053      -2.151       0.541
       intended to comply?    -1.476     0.229     0.866 -1.704   0.088     3.500      -3.173       0.222
       satisfied with service -0.517     0.596     0.533 -0.970   0.332     1.590      -1.561       0.528
       camera                  1.463     4.317     0.755  1.937   0.053     4.246      -0.017       2.942
       _intercept              1.534     4.635     1.296  1.184   0.237     2.080      -1.006       4.073
sigma_ _intercept              0.554     1.739     0.115  4.798 <0.0005    19.249       0.327       0.780
---
Concordance = 0.679
Log-likelihood ratio test = 15.575 on 6 df, -log2(p)=5.946
```

The difference in partial log-likelihood (LL) = -0,364

Log-Likelihood Ratio Statistic (LRS) = 0,728

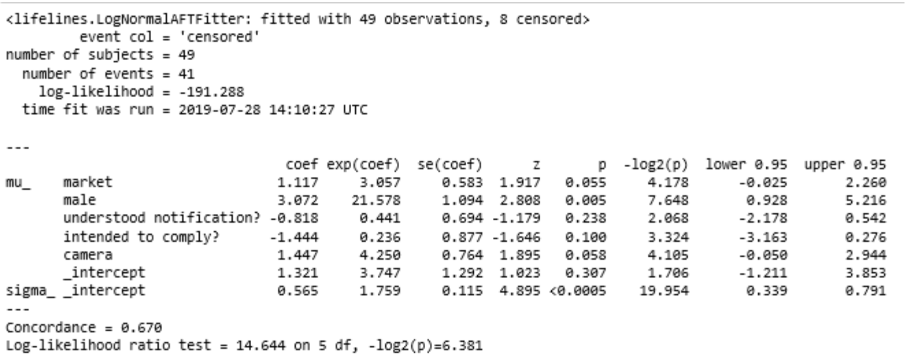
Consulting the chi-square distribution for 0,728 on 1 df: p>0,25

Model 9 is better than model 8

Step 10

The ‘satisfied with service’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, male, understood notification, intended to comply, camera]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -191.288
  time fit was run = 2019-07-28 14:10:27 UTC

---
                              coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market                1.117     3.057     0.583  1.917   0.055     4.178      -0.025       2.260
       male                  3.072    21.578     1.094  2.808   0.005     7.648       0.928       5.216
       understood notification? -0.818  0.441     0.694 -1.179   0.238     2.068      -2.178       0.542
       intended to comply?  -1.444     0.236     0.877 -1.646   0.100     3.324      -3.163       0.276
       camera                1.447     4.250     0.764  1.895   0.058     4.105      -0.050       2.944
       _intercept            1.321     3.747     1.292  1.023   0.307     1.706      -1.211       3.853
sigma_ _intercept            0.565     1.759     0.115  4.895 <0.0005    19.954       0.339       0.791
---
Concordance = 0.670
Log-likelihood ratio test = 14.644 on 5 df, -log2(p)=6.381
```

The difference in partial log-likelihood (LL) = -0,466

Log-Likelihood Ratio Statistic (LRS) = 0,932

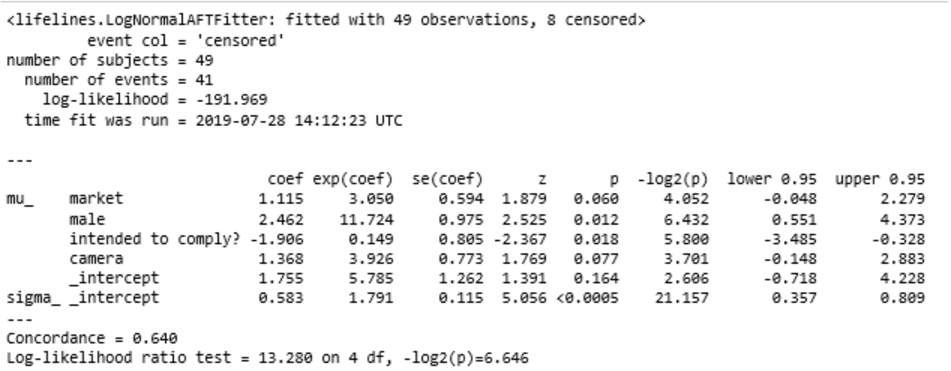
Consulting the chi-square distribution for 0,932 on 1 df: p>0,25

Model 10 is better than model 9

Step 11

The ‘understood notification’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, male, intended to comply, camera]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -191.969
  time fit was run = 2019-07-28 14:12:23 UTC

---
                         coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market           1.115     3.050     0.594  1.879   0.060     4.052      -0.048       2.279
       male             2.462    11.724     0.975  2.525   0.012     6.432       0.551       4.373
       intended to comply? -1.906  0.149     0.805 -2.367   0.018     5.800      -3.485      -0.328
       camera           1.368     3.926     0.773  1.769   0.077     3.701      -0.148       2.883
       _intercept       1.755     5.785     1.262  1.391   0.164     2.606      -0.718       4.228
sigma_ _intercept       0.583     1.791     0.115  5.056 <0.0005    21.157       0.357       0.809
---
Concordance = 0.640
Log-likelihood ratio test = 13.280 on 4 df, -log2(p)=6.646
```

The difference in partial log-likelihood (LL) = -0,681

Log-Likelihood Ratio Statistic (LRS) = 1,362

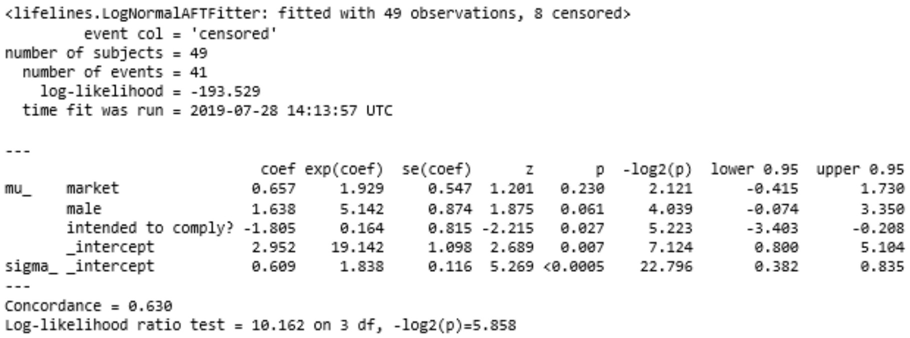
Consulting the chi-square distribution for 1,362 on 1 df: p>0,10

Model 11 is better than model 10

Step 12

The ‘camera’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, male, intended to comply]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -193.529
  time fit was run = 2019-07-28 14:13:57 UTC

---
                         coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    market           0.657     1.929     0.547  1.201   0.230     2.121      -0.415       1.730
       male             1.638     5.142     0.874  1.875   0.061     4.039      -0.074       3.350
       intended to comply? -1.805  0.164     0.815 -2.215   0.027     5.223      -3.403      -0.208
       _intercept       2.952    19.142     1.098  2.689   0.007     7.124       0.800       5.104
sigma_ _intercept       0.609     1.838     0.116  5.269 <0.0005    22.796       0.382       0.835
---
Concordance = 0.630
Log-likelihood ratio test = 10.162 on 3 df, -log2(p)=5.858
```

The difference in partial log-likelihood (LL) = -1,56

Log-Likelihood Ratio Statistic (LRS) = 3,12

<table>
<tr><td></td><td>

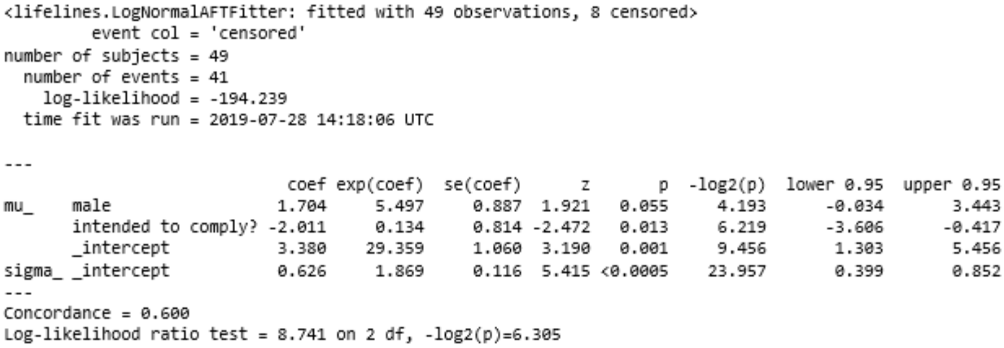
Consulting the chi-square distribution for 3,12 on 1 df: p>0,05

Model 12 is better than model 11

</td></tr>
<tr><td>Step 13</td><td>

The ‘market’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, male, intended to comply]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -194.239
  time fit was run = 2019-07-28 14:18:06 UTC

---
                             coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    male                 1.704     5.497     0.887  1.921   0.055     4.193      -0.034       3.443
       intended to comply? -2.011     0.134     0.814 -2.472   0.013     6.219      -3.606      -0.417
       _intercept           3.380    29.359     1.060  3.190   0.001     9.456       1.303       5.456
sigma_ _intercept           0.626     1.869     0.116  5.415 <0.0005    23.957       0.399       0.852
---
Concordance = 0.600
Log-likelihood ratio test = 8.741 on 2 df, -log2(p)=6.305
```

The difference in partial log-likelihood (LL) = -0,71

Log-Likelihood Ratio Statistic (LRS) = 1,42

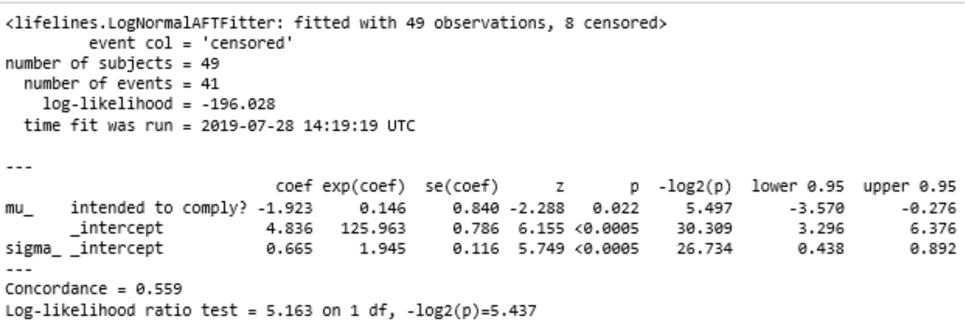
Consulting the chi-square distribution for 1,42 on 1 df: p>0,10

Model 13 is better than model 12

</td></tr>
<tr><td>Step 14</td><td>

The ‘male’ variable is the least reliable and excluded

Modeling the AFT LogNormal model for [market, male, intended to comply]



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -196.028
  time fit was run = 2019-07-28 14:19:19 UTC

---
                             coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    intended to comply? -1.923     0.146     0.840 -2.288   0.022     5.497      -3.570      -0.276
       _intercept           4.836   125.963     0.786  6.155 <0.0005    30.309       3.296       6.376
sigma_ _intercept           0.665     1.945     0.116  5.749 <0.0005    26.734       0.438       0.892
---
Concordance = 0.559
Log-likelihood ratio test = 5.163 on 1 df, -log2(p)=5.437
```

The difference in partial log-likelihood (LL) = -1,789

Log-Likelihood Ratio Statistic (LRS) = 3,578

Consulting the chi-square distribution for 3,578 on 1 df: p>0,10

Model 14 is better than model 13

</td></tr>
<tr><td>Step 15</td><td>

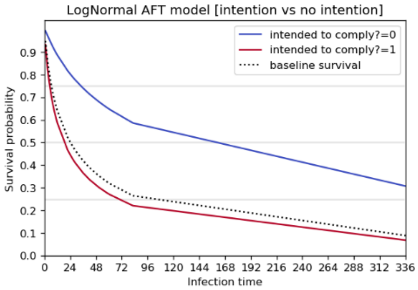
Model 14 is best of all models.

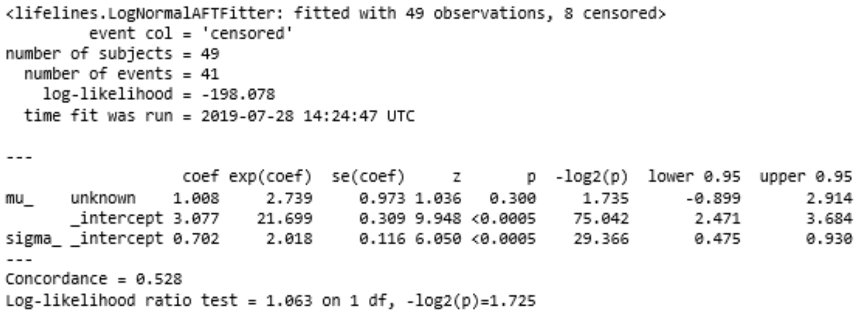
When comparing model 14 with a trivial model, the LRS is 5,163 for 1 degree of freedom: p<0,05

Model 14 is better than a model without covariates.

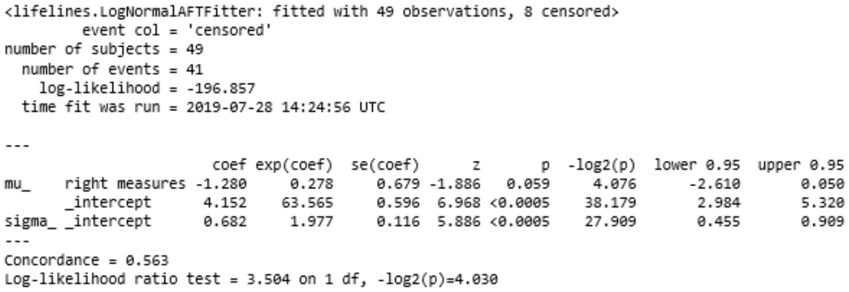
Model 14 is accepted

</td></tr>
</table>

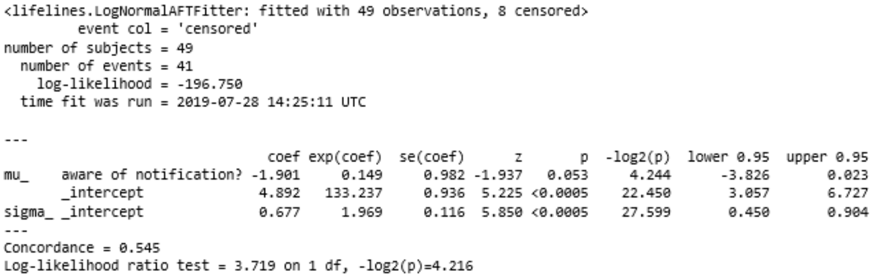




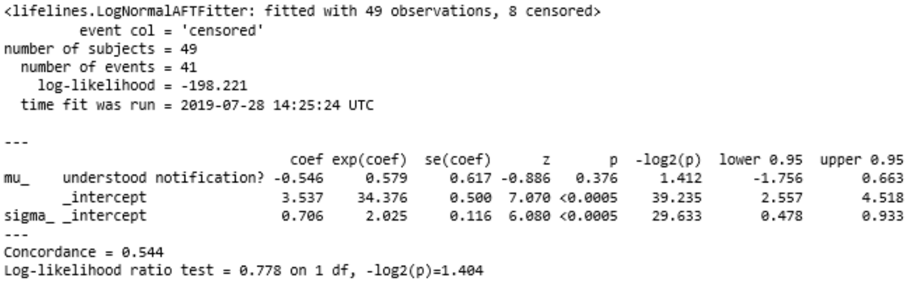
```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -198.078
  time fit was run = 2019-07-28 14:24:47 UTC

---
                   coef exp(coef)  se(coef)     z       p  -log2(p)  lower 0.95  upper 0.95
mu_    unknown    1.008     2.739     0.973 1.036   0.300     1.735      -0.899       2.914
       _intercept 3.077    21.699     0.309 9.948 <0.0005    75.042       2.471       3.684
sigma_ _intercept 0.702     2.018     0.116 6.050 <0.0005    29.366       0.475       0.930
---
Concordance = 0.528
Log-likelihood ratio test = 1.063 on 1 df, -log2(p)=1.725
```



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -196.857
  time fit was run = 2019-07-28 14:24:56 UTC

---
                         coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    right measures -1.280     0.278     0.679 -1.886   0.059     4.076      -2.610       0.050
       _intercept      4.152    63.565     0.596  6.968 <0.0005    38.179       2.984       5.320
sigma_ _intercept      0.682     1.977     0.116  5.886 <0.0005    27.909       0.455       0.909
---
Concordance = 0.563
Log-likelihood ratio test = 3.504 on 1 df, -log2(p)=4.030
```



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -196.750
  time fit was run = 2019-07-28 14:25:11 UTC

---
                                 coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    aware of notification? -1.901     0.149     0.982 -1.937   0.053     4.244      -3.826       0.023
       _intercept              4.892   133.237     0.936  5.225 <0.0005    22.450       3.057       6.727
sigma_ _intercept              0.677     1.969     0.116  5.850 <0.0005    27.599       0.450       0.904
---
Concordance = 0.545
Log-likelihood ratio test = 3.719 on 1 df, -log2(p)=4.216
```



```
<lifelines.LogNormalAFTFitter: fitted with 49 observations, 8 censored>
         event col = 'censored'
number of subjects = 49
  number of events = 41
    log-likelihood = -198.221
  time fit was run = 2019-07-28 14:25:24 UTC

---
                                    coef exp(coef)  se(coef)      z       p  -log2(p)  lower 0.95  upper 0.95
mu_    understood notification? -0.546     0.579     0.617 -0.886   0.376     1.412      -1.756       0.663
       _intercept                3.537    34.376     0.500  7.070 <0.0005    39.235       2.557       4.518
sigma_ _intercept                0.706     2.025     0.116  6.080 <0.0005    29.633       0.478       0.933
---
Concordance = 0.544
Log-likelihood ratio test = 0.778 on 1 df, -log2(p)=1.404
```

All models with a substitution for the variable intention are not accepted because the LRS for 1 degree of freedom is for all lower than 3,841 (critical value for p-value of 0,05).

Model 14 is thus the best model.

TUDelft

kpn