TALLINN UNIVERSITY OF TECHNOLOGY

School of Business and Governance

Department of Business Administration

Tuomas Tiainen

Customers' awareness of data collection in online marketing

Bachelor's thesis

Programme TVTB, specialisation Marketing

Supervisor: Andrei Spiljov

Tallinn 2018

I declare that I have compiled the paper independently
and all works, important standpoints and data by other authors
have been properly referenced and the same paper
has not been previously been presented for grading.
The document length is 7873 words from the introduction to the end of summary.

Tuomas Tiainen ……
(signature, date)
Student code: 131087
Student e-mail address: tuomastiainen3@gmail.com

Supervisor: Andrei Spiljov:
The paper conforms to requirements in force

……
(signature, date)

Co-supervisor:
The paper conforms to requirements in force
……
(signature, date)

Chairman of the Defence Committee: / to be added only for graduation theses /
Permitted to the defence
……
(name, signature, date)

TABLE OF CONTENTS

ABSTRACT

For this bachelor's thesis, an online survey is used to study people's general knowledge about online data collection, their confidence in these practises and also how much people trust corporations to keep the collected data safe. The research focuses on studying online data collecting principles, methods and how the collected data is used for targeted marketing. Side focus is on investigating how common this activity has become and what it means to people's privacy and online security. Another essential objective of this research is to examine how well people know that many companies and service providers collect user data online for advertising purposes and therefore, how these practises affect people's trust in companies.

The research consists of two research methods, qualitative and quantitative. Observation was used as a qualitative method when studying different previously published articles and studies close to the topic of the thesis. For the quantitative method, an online survey is used to examine how well people know data collection and how it affects their trust in the companies.

The overall findings show that people are aware that companies and service providers collect user information for marketing purposes, but most people do not accept the fact that the collected information might be exposed to third parties. Results also show that people are not sure what websites´ privacy policies mean in regard to sharing collected data with other parties. Furthermore, findings show that people's trust in companies will fall if collected user data is exposed to third parties.

As a recommendation for online companies, customers should be provided with better knowledge of online data collection and what concerns it has for people's privacy. Therefore, companies should provide a clear overview of what data they are collecting and for what reasons. Companies and service providers should also make sure that user information doesn´t get into the hands of third parties without users permission, because customer´s trust in companies will most likely decline.

Keywords: Online marketing, Target marketing, Data collection, Online privacy, Web cookies

INTRODUCTION

Effective online marketing depends increasingly on the use of customer data collected from the internet. With the explosion of digital technologies, companies and website owners can collect very detailed and extensive information about people's activities online. Online data collection has become a very important business for the companies which are offering sophisticated marketing tools for their clients.

For any company, collecting information about their customers helps companies to have a better understanding of their customers' needs and wants. This kind of activity is no different from the basic idea of marketing: trying to understand the customer. The main goal is to collect detailed customer information and then use it as effectively as possible in many marketing practises. Before, collecting customer information was highly visible (surveys, polls, etc.), with the well-defined samples. Nowadays when people may have their whole life in the internet, while many companies have their activities in there, collecting information has become invisible, more detailed and even larger samples can be collected. This means that more and more advertising can be targeted to right target groups.

However, with large amounts of data, companies have a huge responsibility to keep it safe and make sure that it doesn't get exposed to third parties. Therefore, companies should understand what feelings the almost invisible data collection is causing in customers and people in general, and how it might affect people's trust in the companies. This bachelor´s thesis observes and studies general knowledge of online marketing and how data is collected for marketing purposes. The objective of this research is to study how much people know about online data collection and how it affects trust in companies which are doing that.

The main research questions are:
1) What is people's general knowledge about data collection?
2) How does online data collection affect people's trust in companies?
3) Do people trust in companies´ ability to maintain customer information safely?

Data for this study is collected through a survey and it is conducted by asking questions related to general knowledge of data collection and how people are relying on companies to handle information confidentially. The study will focus on people who are using the internet and their knowledge of online data collection methods in general and how much they trust in companies doing so. Side focus is to understand how data is collected, which tools are used, how common this activity has become and what does it mean to people's privacy and online security. Therefore, this study will provide more perspective to online marketing and interesting information about it.

Usually studies in this area are lacking deeper analysis of how data collection affects the customers´ trust. Fundamentally it is a question concerning the incentives of a company's decisions to be open about their data practices. Even though many organizations are open about their data collection practices and obey laws, others prefer to keep consumers in the dark. Also, it is not unusual for companies to quietly collect personal data they have no immediate use for, reasoning that it might be valuable at a later stage in the future. The ultimate goal is to raise more discussion about online data collection and how it affects people's privacy and trust in website providers which collect user information for their marketing purposes, and how these providers handle huge amounts of customer information confidently.

In order to achieve its objective, the thesis is divided into three parts: theoretical framework, methodology of the research methods and results, and discussion. The literature research in the first and second chapter has been collected from relevant literature, secondary data and articles. This acts as a theoretical basis for the whole study. The first two parts provide an overview of online marketing, data driven marketing, data collection tools, and the prevalence of data collection. Additionally, the research will take a detailed look into how the collected data can be used in different marketing solutions. The third part introduces the reasoning behind the choosing of this topic, research methods and results. The last part presents the research analyses. The topic for this thesis was generated from the author`s interest towards data driven marketing opportunities online, as well as the desire to study people´s trust in companies that are using data collection.

1. THEORETICAL FRAMEWORK

All marketing disciplines are driven by some type of data, therefore marketers take all the available information into account. Online marketing has the capacity to remember a customers´ preferences in great detail. The evolution of the internet has caused more and more people to spend most of their time online. People share their lives in various social networking services, which means that there is a huge amount of data created every day. In different social media platforms, people share their interests in gourmet food or football, for example. This kind of data is a gold mine for the marketers.

Online marketing, data collection, web cookies, advantages, and disadvantages are introduced as a basis of theoretical framework in the following chapter.

1.1 Online marketing

There are many terms for marketing activities that happen on the internet. The using of these terms might vary by location and what different marketing networks are included in the process. The terms online marketing, internet marketing and E-marketing, are frequently exchanged, and can often be considered synonymous. Marketing itself is typically defined as an action or business of promoting and selling products or services, including market research, advertising, and delivering products to people.

Online marketing doesn’t fundamentally differ from marketing - the difference lies in the environment where marketing actions happen. Awad (2006) emphasizes the role of the internet in connecting people and processes to suppliers, customers, and business partners. A process means reaching people to consummate a transaction or to deliver product. There are also other definitions which include using different online channels for marketing. According to Awad, internet marketing or online marketing refers to advertising and marketing efforts that use the web and email to drive direct sales via electronic commerce, in addition to sales leads from websites or emails. In this case marketing actions are made via email and the web.

These definitions strongly highlight the internet as a marketing platform, but it is important to keep in mind that there are also many other channels in marketing. According to Charlesworth (2007), using terms online marketing and E-marketing both refer to any aspect of the discipline of

marketing that is performed in an internet environment. However, under the term E- marketing, we could include many other marketing channels such as wireless media, mobile networks and radio. E-marketing uses a range of technologies to help connect businesses to their customers, so the term online marketing tends to refer more to marketing products and services over the internet and alternative term E-marketing can considered to have a broader scope since it refers to any use of technology to achieve marketing objectives (Chaffey 2004).

For this bachelor´s thesis, the author prefers to use the term "online marketing" because it emphasizes more the use of the internet in marketing actions.

1.2 The importance of online marketing

The requirements and wants of the consumers can be better understood by knowing more about the people and also analyzing ways of providing for their needs. Online marketing enables resources which are different kinds of tools to complete the marketing actions. Typical technology behind marketing resources involve the internet and interactive websites. These resources are primary related to communications, data processing and trading transactions (Chron 2018).

The internet allows companies to stablish deeper relationship with their customers in a way that maintaining customer is more likely. Social media marketing allows companies to share company announcements and keep their customers up to date with a new product or service features. The internet makes possible to create personalized marketing content by using different data collection methods. Interactive websites can be personalized for particularly user by offering content which is based on user's habits and purchase history. These data collection methods also allow companies to gather necessary information, which will improve companies decision making processes. Online marketing enables cost effective marketing methods compared to traditional marketing. Different types of online marketing strategies offer efficient ways for companies to take full advantage of their marketing efforts, and boosting return on their investment (Optimus01 2017).

These communications, transactions activities and information research are all parts of traditional marketing but marketers can improve these traditional marketing processes by using more efficient tools which online marketing has to offer.

1.3 Online tracking

Every single move made online can be, and very often is, tracked by website owners and advertising networks that gather user information for using it for personalizing user experience and last mentioned for targeted advertisements. As users browse and interact with websites, they are observed by both first-parties, which are the sites the user visits directly, and third-parties which are typically invisibly trackers such as advertising networks behind most websites. Usually website owners perform first-party tracking to personalise user experience across web sessions, but it is also used for fraud detection and law enforcement (Casteluccia 2012). Third-party tracking can obtain users' browsing histories through a combination of HTTP cookies and other tracking technologies that allow them to uniquely identify users, and mark them way which tells the third party which first-party site the user is currently visiting. Third-party tracking makes possible user profiling which can be used for targeted advertising and web analytics/measurement (Casteluccia 2012). Many commercial websites uses online tracking for these reasons in first- party contents but also third- parties use these same methods.

1.3.1 User profiling and behavioural advertising

User profiling means that online advertisers collect several pieces of data about web user, which revels their browsing behaviour. Data includes information such as; how many pages browsed on a website, the time spent on the website, how many clicks made by user, and the overall interaction with the website (Kissmetrics 2014). Purpose of user profiling is to build user profiles which have information about their interests, main characteristics such as age and gender, and shopping habits (Casteluccia 2012). These user profiles are used for behavioural advertising, which helps online advertisers to present targeted advertising to users that reflect their interests making advertising more efficient (Kissmetrics 2014). Therefore, online user data is collected and used to target users with more relevant advertising content.

1.3.2 Web analytics/measurement

Online tracking is not used only for behavioural advertising but also for measuring different aspects of website usage. Web analyses is a tool which purpose is to understand behaviour of the website visitor and use that information to optimize website content and measure effectiveness of advertising campaigns. Many websites has a third- party presence because they are using third-party tools such as Google analytics (Casteluccia 2012). This third- party presence is discussed in more detail in the next part.

The website which is using Google analytics, has a tracking code which is downloaded to user's computer every time user visits the website. The code captures data which has an information about what website user is viewing and title of the page, what web browser user is using, and user's location and what language is used (Rosche 2016). This data is used for optimizing and personalizing web content for the user, example online retail website shows automatically similar products what user is purchased in the past.

1.4 The most popular tools used for tracking users

Online marketing cannot function without data collected from the users. The data is analysed and used for user profiles, which help online marketers to target relevant advertises to a people. The main source of information comes from different web tracking technologies. This is the important difference between ordinary marketing and online marketing. Website users generate digital trails and data that may be stored and analysed by the website owner operating the web server. That is why there are different essential tools that website owners are using for tracking.

1.4.1 HTTP Cookies

Web cookies, or HTTP cookies, are small files which are stored on a user's computer. They are designed to hold a small amount of data specific to a particular user and website. Cookies are used for session management, personalization and tracking (MND web docs 2018). Each time a user visits a new website, cookies are created by browser and when that user returns to the website, the

cookies will help it to remember certain things, such as what content the user viewed and which pages they accessed.

For online marketing, cookies play important role in which type of advertising content is shown to people. Cookies track what websites user has visited and what products are browsed, and this information helps online marketers to target advertises related to user's interest and browsing history (Markelz 2016). Therefore, if someone has browsed kitchen equipment in one site and user moves to the other site. User is served advertises about products which they just viewed.

Difference between cookies comes from who has created them. Cookies have a domain associated to them, so if a cookie´s domain is the same as the domain of the website where user is currently on, the cookies are first-party associated. When the domain is different, then it's called a third-party cookie (MND web docs 2018).

1.5.2 First party cookies

In a first-party context, cookies help website owners to serve more meaningful content and understand better how users are using their websites. First-party cookies are website related, meaning that they are coming directly from the websites which user has visited (Zawadzinski 2016). Most of the first-party cookies are designed to keep track of user movements within the site, help user resume where left off, remember users registered login, site language, preferences, and other customization functions. The website stores a corresponding file (with same ID tag) to the one they set in users browser and with this file the website can track and keep information on users movements within the site and any information user may have voluntarily given while visiting the website, such as email address. These unique identifiers are first-party data which help website owners to identify visitors and follow their behaviour (Markelz 2016).

1.5.3 Third-party cookies

While many websites just use cookies for session management and personalization based on the user's preferences, many commercial websites use a third- party cookies. These type of cookies work similar than first-party cookies, except they are created by other parties, not the website user visits (Zawadzinski 2016). Many websites include advertising material which is coming from a

third-party site, and it's possible that those advertises could store their cookies to users' computer (Capitol media solutions 2017). As said earlier, third-party cookies are collecting same type of information than first-party cookies, such as the user's interests, location, age, the name of the site, and websites visited. Third-party cookies can also track a user's behavior, such as the content they view on the website and what kind of products they search. When user later visits the website, which include same kind of third-party advertises, the advertiser will be able to read the information, and can obtain user's browsing histories. This allows advertisers show advertises based on the user's past activity and browsing history. Third-party cookies are commonly used, therefore many website has a third- party presence.

Third-party cookies are really important for online advertising but there are many problems related to them. Using third-party cookies might cause privacy issues and if website doesn't state clearly in their privacy policy what cookies site uses, and cookie use is discovered, this might decrease peoples trust (MND web docs 2018).

2. THE THIRD-PARTY DATA BREACH PROBLEM

All the websites are using web cookies but many of them are also using different tracking tools owned by third parties which collect user's private information or unique identifiers. Many of websites leaks this information- often in purpose to third-party tracking sites but sometimes this behavior is unintentionally. In this part of thesis, focus is on to understand how widespread this problem is and how vast it has become. The author of thesis has chosen different studies to illuminate how many third parties can be found in websites.

2.1 Third-party presences in websites

Everyone who uses internet on daily basis, visits different types of websites hundreds of times per month. Some of the websites are hugely popular, some of them less popular. As said earlier, different third-party trackers are important for online advertising. Thus it is important to find out how many third- party trackers can be found in different websites.

Research made by Englehardt and Narayanan (2016), shows that when measuring how many third-party trackers can be found in top 1 million websites, they found out that the total number of third parties present on at least two first parties out of 1 million is over 81,000 (Engelhardt and Narayanan 2016). The major finding was that generality of third parties quickly drops off so that only 123 of these 81,000 are present on more than 1% of sites. Conclusion was that only popular commercial websites have a third-party presence and that there is still only a small probability for a user to encounter same third parties on a daily basis (Engelhardt and Narayanan 2016). But who owns most of these third-party trackers? In the previous study and also a study made in 2014, these owners are found. According Falahrastegar, Haddadi, Uhlig and Mortier (2014), while examining how vast third-party tracker practises has become globally, companies such as Google, AOL and Yahoo appear to own a large number of third-party trackers. Both of these studies included the same companies. This is not a surprise, because these companies are search engines which offer online advertising services for their customers and provide tracking tools for collecting large amounts of user data for targeted advertising. This means that a handful of companies have an enormous amount of data about browsing habits and interests. According to Lewis (2017), 45% out of the 1000 most popular websites on the internet actually use same tracking tools. The

conclusion of Lewis´ study is that these tracking tools are communicating with each other and therefore could share user information.
To sum it all up, all of these studies show clearly that almost every popular website has a third-party presence and only a handful of companies own these third-party trackers.

2.2 Information shared from website to website

A common thought might be that only less popular websites leak personal information purposefully to third-party sites that track users´ browsing behavior for advertisers. But this type of thought is proved wrong, as noted in previous section. Many popular websites have a third-party presence, because almost every website is using the same tracking tools owned by only a handful of companies. These tracking tools can then share information with each other and in that way connected to each other.

Research made by Wills (2011) explored websites that encouraged users to register themselves on popular travel and health sites. These sites were chosen because in these websites, during the registering process, the user has to share private information such as their name and address, and on health and travel sites user searches can reveal one's health issues and travel plans. Research found out that three quarters of 100 popular websites used by tens of millions of people directly share either private information or users unique identifiers to third-party tracking sites, often on purpose. But it is not always a decision made by website providers, whether private information is shared deliberately or intentionally, tracking cookies might still share collected user information with each other.

Cookie syncing is a process (also explained and studied in the study of "A 1-million-site Measurement and Analysis" made by Englehardt and Narayanan) where users´ unique identifiers are scanned from one website to another. This allows cookies from the same origin to share users´ unique identifiers with each other. Therefore, cookie syncing enables advertisers to provide online advertisements for right target audiences (Zawadzinski 2016). Combining the results of these studies, the outcome is that most of the private information and users´ unique identifiers are most likely shared to different websites without the users' permission. This is mostly because third-party trackers which are from same origin can share user data with each other. That is how search

engine companies like Google can offer efficient marketing services for their customers. As a conclusion, cookie syncing means that even the websites´ owners aren't always sure what purposes the data of their customers is used for.

2.3 General views about data collection and privacy

Public opinion about privacy is that it is really important for people - even though many feel that it is not possible to stay anonymous online. A survey conducted by the Pew Research Center investigated how the majority of adults feel about their personal privacy offline and online. The survey revealed that most adults have a desire to improve their online security, though many believe that they cannot achieve anonymity online easily (Madden 2014).

Another survey conducted by the Pew Research Center studied Americans´ views about data collection and security. Participants were asked about their daily interactions online/offline, and the result was that many of the privacy-related values are deeply important for people, and especially having a sense of control over who collects information and how these activities can be observed (Madden and Rainie 2015). A particularly interesting finding were participants´ feelings towards different types of organizations that retain collected information. The result was that people are less comfortable with certain online service providers, such as search engine providers and social media sites, which store user information. Especially online advertisers who place ads on the websites that the customers visit, should not save any information about their activity (Madden and Rainie 2015).

Another interesting topic is how familiar people are with privacy policies that service providers are announcing on websites. A survey conducted by Software Advice examined how many users actually read these documents all the way through before agreeing to their terms, and the finding was that only 8% of people always read them. Almost half of the participants never read these documents (Humbries 2014). The same study also investigated the public´s opinions on data collection and people's attitudes towards companies that also track users´ external web activity. Opinions were mixed about data collection. Less than half of the respondents felt that their privacy deteriorated, while the rest of the respondents did not have an opinion or saw that the loss of

privacy is the cost of using the internet (Humbries 2014). When asked about companies acting as a third party and collecting user's external web activity, majority of the respondents saw this activity as unacceptable.

To summarize these results, many people have serious concerns about their privacy and security online, but many think that they cannot do anything about it. People have knowledge of which types of service providers might collect data, and how advertisers collect user information for targeted advertising. Many have bad thoughts about the activities of these third parties, especially those of online advertisers. However, people do not bother to explore websites´ privacy policy. These previous studies give a very comprehensive picture of the people's thoughts about data collection, but not how these thoughts affect people´s trust in internet service providers.

3. METHODOLOGY OF THE RESEARCH AND RESULTS

This chapter focuses on research and includes an introduction for subject choice, an overview of the research design, and explains the data collection methods used.

3.1 Introduction for Subject Choice

People use internet on a daily basis and visit several websites. Almost every website has a notification of the use of HTTP cookies. In fact, several regulations require that website providers have to give clear information about HTTP cookies and how the collected data is used. Practices vary from one country to another but for example the Finnish law states that users must be given clear and complete information on cookies and the storage and use of data concerning the use of the service. Service users must be asked for approval for saving and using their data. Websites should present information and allow users to deny the storage of data in the most user-friendly manner possible (Finlex 2011). Even though many people see these notifications, most of them don't bother reading websites´ privacy policies, and that way understand how information of their actions and behaviour is used (Humbries 2014). Therefore, it would be important to study how many people actually understand how HTTP cookies can be used and what it might mean to their privacy.

A research made by Turow, Mulligan and Hoofnagle (2007) studied consumers´ beliefs about the term "privacy policy". They believe that it creates substantive rules limiting the collecting and use of data, even though data collection is common in online marketing and targeted advertising. The research reveals that consumers do not understand the nature and legality of information-collection techniques that form the core of online advertising. This study inspired the author's interest about the topic and it also underlined the importance of finding out how well people know HTTP cookies. First party cookies work similarly to third party cookies. The difference between them lies in who has created them. The author assumes that many people don´t acknowledge how many websites might have third-party presence and how websites can collect private information and unique identifiers of users.

As noted in the theory chapter of this research, a handful of corporations own the majority of all third-party cookies. That is because these corporations need to collect huge amounts of user data in order to provide different marketing tools for their customers. Perhaps the biggest of them all is Google, which offers many different services and solutions for advertising and marketing. For advertising, Google offers the following solutions:

- Search ads: advertisements will appear on Google search results when users search for products and services that are closely related to the products and services of the advertiser. Customer only pays to Google when users visit the website or make a call by clicking on ad (Google 2017).
- Display ads: Display ads are displayed on more than two million websites and over 650,000 apps. With display ads, accurately targeted campaigns can be created based on customer information, such as interest or demographic information (Google 2017).
- Video ads: Video ads are only shown to audiences that are most likely to be interested in the product or service. Audience can be chosen by age, gender, location, or interest. So for example, advertising can be targeted to sports enthusiasts, music enthusiasts, or any other target group (Google 2017).
- App promotion: apps are promoted to iOS or Android users on Google. Once again, ads are optimized so that they reach the right target audience (Google 2017).

For analysing website traffic, Google provides a service called "Google Analytics". It is used for tracking and reporting how users behave on the website, and finding out the ROI for online marketing. Google Analytics tracks all of its data by adding a unique tracking code to every page of a website, and also a cookie on user´s computer, which provides anonymous information to create unique identifiers of the users (Google 2017).

This is a prime example of how these marketing solutions and tools can only work if enough data is collected from the users of the internet. Because several previously made researches show that many third-party cookies are owned by Google, we could assume that many website owners are using Google's services. Thus it is important to understand how widespread and how common this activity has become, and it is really important to study how well different website visitors understand the situation where their every move on the internet might be tracked for marketing purposes.

Studies show that privacy-related values are deeply important for people and especially having a sense of control over who collects information and how these activities can be observed (Madden and Rainie 2015). Therefore, it would be important for companies and website providers to understand how much data collection effects people's trust and how much people rely on them to keep data out of third parties.
The objective of this study is to research how good internet users´ general knowledge of online data collection is and how much users trust service providers to handle their data with care.

3.2 Research design overview

In order to answer the research questions, the author is using qualitative and quantitative research methods.

The main research questions are:
1) What is people's general knowledge about data collection?
2) How does online data collection affect people's trust in companies?
3) Do people trust in companies´ ability to maintain customer information safely?

Qualitative and quantitative methods were combined to complement each other, based on secondary and primary data respectively. The research is based on customers' knowledge of online data collection, and whether or not they trust that their data is used correctly. Several studies were used for secondary data: these different studies were about how many websites are using data collection technologies, third-party presence in the websites, and consumer beliefs about online privacy. These studies are a good insight into the present state of online marketing and users´ privacy issues, and therefore the studies acted as a stepping stone to the research questions.
A structured online survey was used as a method for data collection.

3.3 Methods

Both qualitative and quantitative research methods were used to answer research questions. The research is based on internet users´ general knowledge of online marketers collecting user data and

how it can be used for advertising without their direct permission, and how trustworthy people see this kind of activity. Author is using a questionnaire to collect data for a quantitative research, therefore the survey can only contain closed or structured questions. This quantitative study is based on measurement values which stem from the collected data. The measurements are then analysed by using statistical analysis methods. Therefore, the mostly used method is quantitative approach of conducting an online survey to internet users, and the survey results will provide numeric data. Possible limitations are related to sample size and that it might include non-representative participants. As for the qualitative methods used, the author has examined different studies of how websites are using HTTP cookies and also previously made studies about people's knowledge and feelings about data collection. This method serves as a basis for building an understanding of the current situation.

To gather data for the research, the author conducted a survey. The aim of the survey was to find out how much people know about data collection and how much people trust in service providers doing so. The first part of the survey focuses on general knowledge of data collection. Respondents were asked about data collection and web cookies. The second part of the survey focuses on participants´ feelings about data collection and how it affects trust in companies. Last part of the survey focuses on finding out how much participants trust in companies to keep collected data safe. The quantitative survey with closed questions was a useful way to reach the objective of the study.

The previous study made by Engelhardt and Narayanan (2016) about third-party presence in different websites acted as an inspiration and pinpointed possible problems about privacy issues. This enabled choosing the research questions and finding out how these data collection actions affect people's trust.

3.4 Results

This chapter will present data that has been collected through a quantitative survey, which was created with SurveyMonkey and respondents been gathered from on the author's workplace. The survey was shared with 124 co-workers and the response rate was 44%. The online survey gained

55 respondents and gender gap was not substantial. Female respondents were 58.18% when male 41.82%. Age of participants ranged from young adults to middle-aged people.

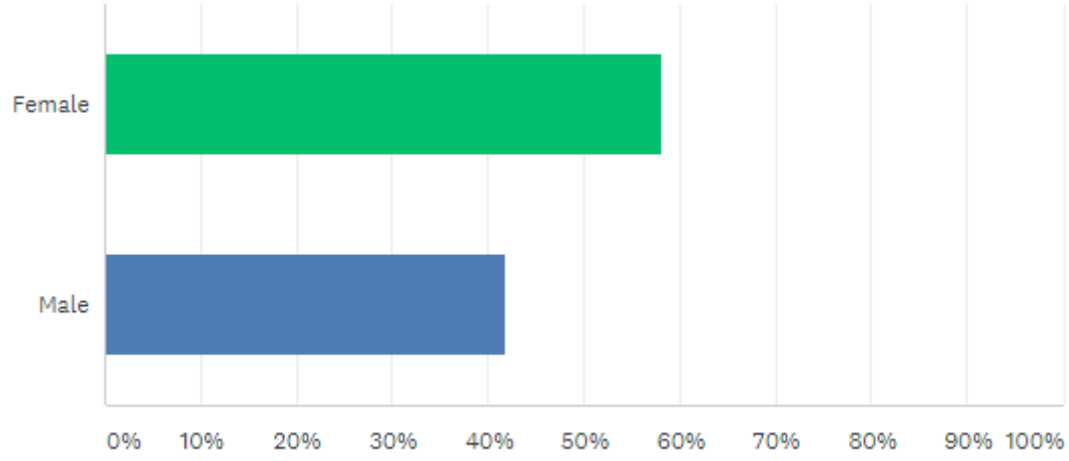

Figure 1. Respondent's gender. Source: Appendix 2, Table 2.1

One of the research questions was to understand how well internet users know that service providers might collect data, example movements at the website, and what they are interested, and how well users know what web cookies are, and what they are used for. Respondents were asked are they aware that many websites might collect that type of data, and 76% were aware of this, 20 % was not aware of this, and 4% respondents answered that they don't care.

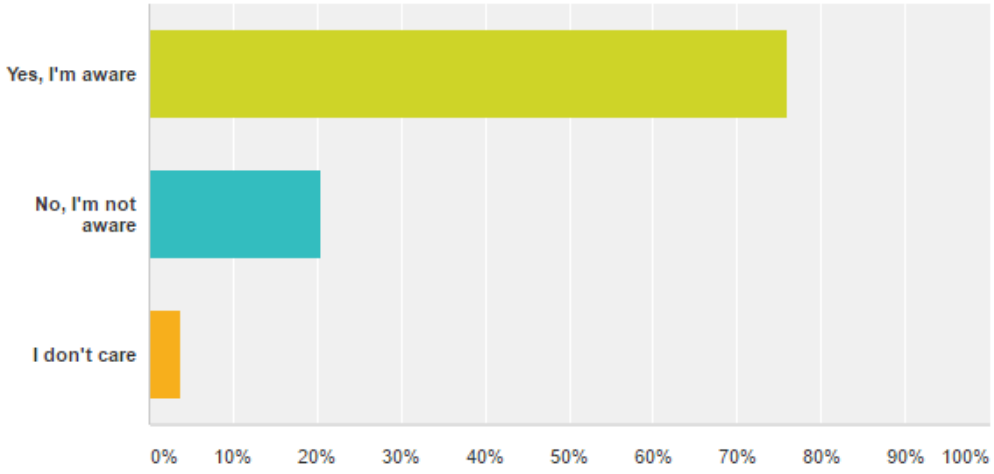

Figure 3 General knowledge about data collection. Source: Appendix 2, Table 2.2

Web cookies play important part for websites functionality and interface experience, but some of cookies are used for collecting user data, which can be exploited for marketing opportunities and user profiling. Web regulations order that service provider has to inform user by showing clear notifications, if using cookies on their website. In figure 4, participants of survey were asked about have they ever paid attention to these notifications, and 89% have paid attention, and 11% have not.

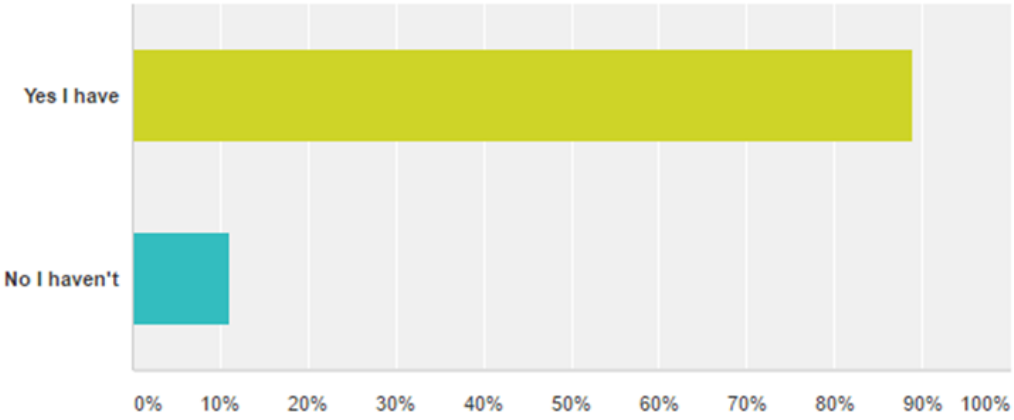

Figure 4 Have you ever paid attention to notifications of firms using cookies on their web site. Source: Appendix 2, Table 2.3

As seen below, participants were also presented a statement, when company has privacy policy, do they know that the site will not share their information with other websites or companies. Result show that 22% believe it is true, 20% said false, and over half of participants 58% answered “I don’t know”.

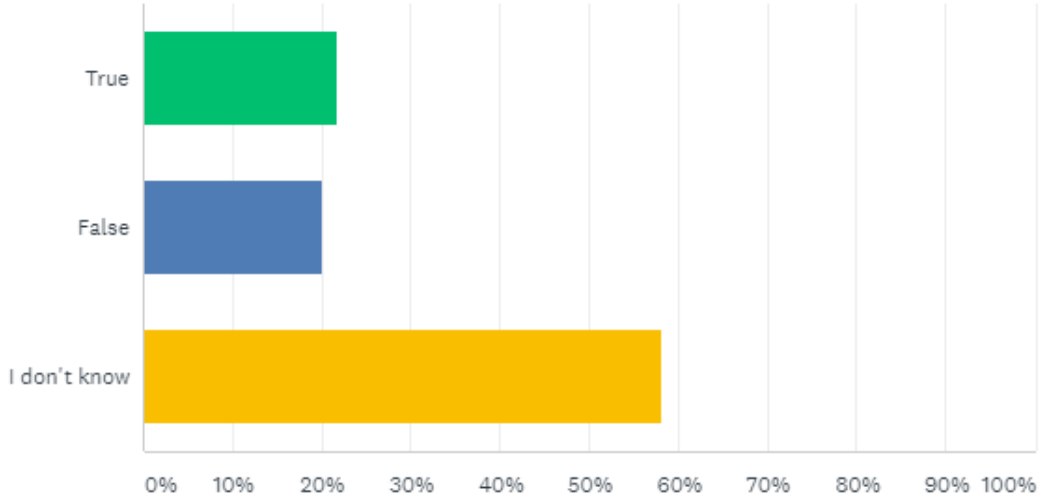

Figure 5 When a web site has a privacy policy, I know that the site will not share my information with other websites or companies. Source: Appendix 2, Table 2.4

The next section of the survey focused on examining how survey participants feel about service providers collecting user data and how it affects their trust in companies. For the first question of this section, participants were asked if they feel that their privacy might be in danger when companies collect data about their web habits, interests and movements. Figure 6 shows that majority of respondents 65% worries that privacy may be in danger, while 25% said that it doesn't worry them, and 9% answered they don't care.

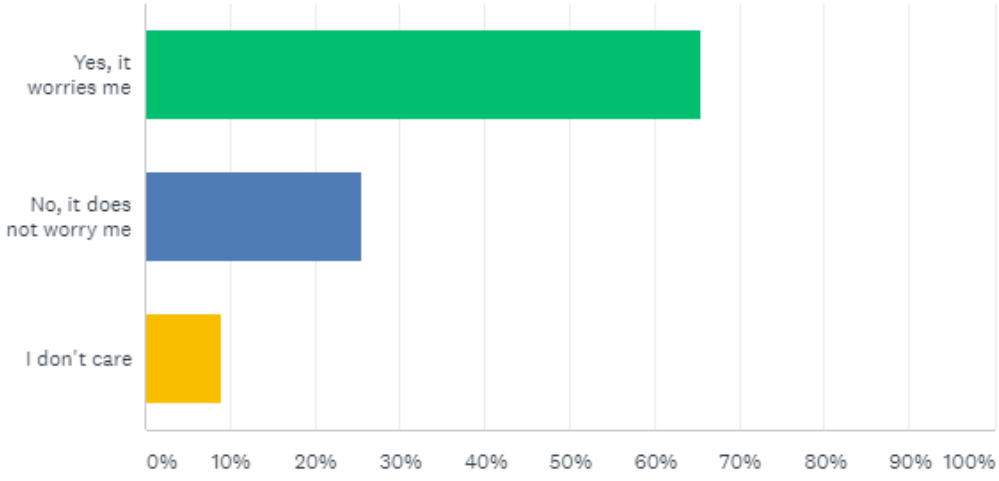

Figure 6 Do you feel that your privacy may be in danger when companies collect information about you and your web habits. Source: Appendix 2, Table 2.5

The next question dealt with fears about if collected user data is exposed to third parties. Many service providers collect large amounts of user data, which can be sold or accidently leaked to bidding corporations. Due to 85% of survey participants feel fear about user data exposed to third parties, while 11% doesn't fear and 4% answered that they don't care.

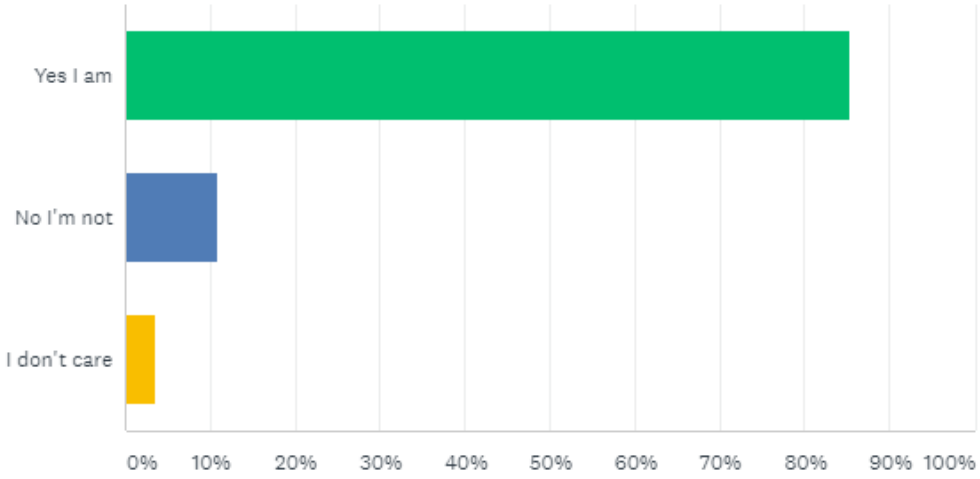

Figure 7 Are you afraid that customer data collected by corporations might be exposed to third parties. Source: Appendix 2, Table 2.6

The last part the of survey aimed to find out how respondents trust on companies which handle large amounts of data, can company handle it confidentially and trust that they can keep it safe. Figure 8 below illustrates how respondents trust in companies to handle collected customer data confidentially. It reveals that the majority of respondents 60% answered that they don't trust, 40% said no, while choice "I don't care" didn't get any responses.

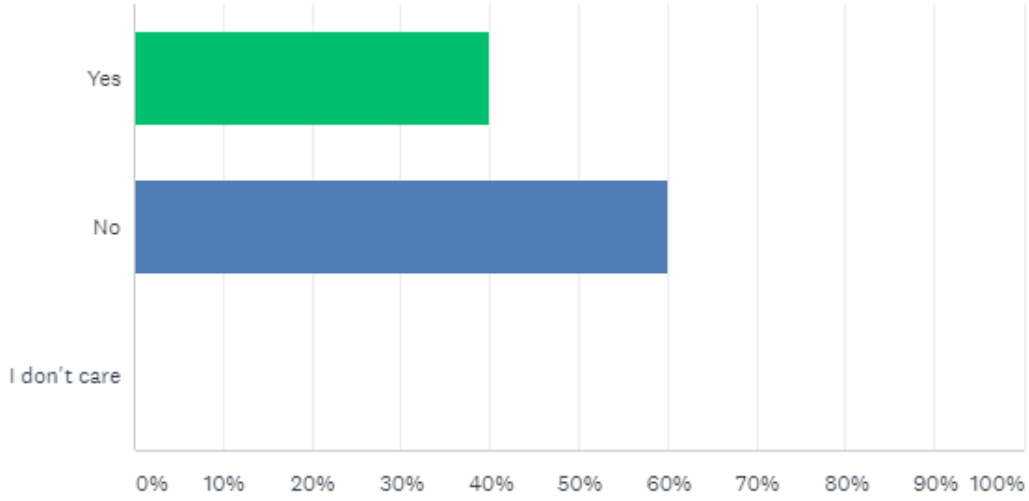

Figure 8 Do you trust companies to handle customer information confidentially. Source: Appendix 2, Table 2.7

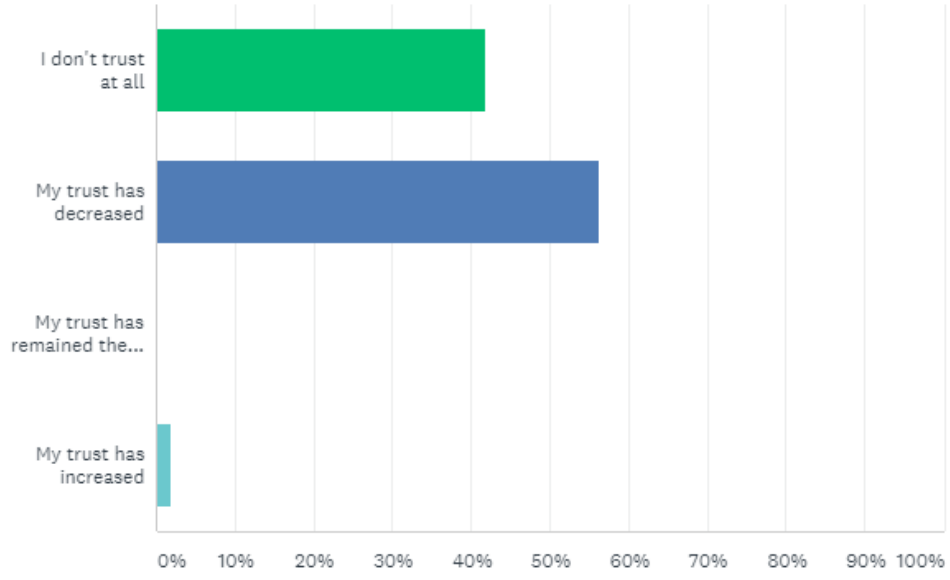

Figure 9 Would you trust a company that has accidentally leaked customer information to a third party. Source: Appendix 2, Table 2.8

Figure 9 demonstrates how respondents trust changes if company accidently leaks customer information to a third party. Most likely, respondents trust decreases 56%, but 42% of respondents doesn't trust at all.

3.5 Findings and recommendations

This chapter will present findings that author has made from observing previous studies, as well as the results gained from the online survey. These findings will be reflected to the theoretical framework presented in chapter 1 and 2. Additionally, recommendations are provided.

The survey was conducted at the author's workplace and there was not a distinct gender gap while participants´ ages ranged from mid-twenties to middle age. Therefore, the results of the survey provide results from a wide range of different age groups. All of the respondents are working in the finance industry where use of internet is an important part of work. That is why the author assumes that many of the survey participants are using the internet on daily basis. Out of 54 respondents, 41 people were aware that every time they visit a website, companies collect information from the users, such as personal interests. This result implies that people understand how marketing works in general and how marketing could not work without customer/user data which is collected for marketing purposes. But if the user data, which is usually detailed information, is in danger end up to a third party, people don´t accept this. In summary, the research results show that data collection is only acceptable if the data will benefit the first party that has collected it to improve their services, not the third parties.

Based on survey results, many participants have noticed notifications about websites using cookies. This indicates that the most popular commercial and company websites obey legislation on cookies, and users are asked if they agree to most cookies and similar technologies before the site starts to use them. These requirements are clearly stated in the EU Commission's guidelines on privacy and data protection and in the ePrivacy directive, specifically Article 5(3). Based on the results of the survey, people were not sure what privacy policies on the websites mean concerning the websites sharing information to other websites. This may mean that very few users are familiar with the websites data protection practices and how online marketing works in general. Results show that online privacy is important for users, and many fear that their privacy is in danger when companies collect private information about them.

The purpose of the last section of the survey was to find out if people trust companies on their data collection methods and if they trust companies to keep their data safe. The results show that many people already believe that collected data is exposed to third parties and many do not trust

companies to keep collected data safe or to handle their information safely. All in all, most people don´t trust companies after the data is leaked to a third party.

The recommendations about spreading knowledge of online marketing and data collection methods are developed based on the conducted survey and observations of previously made studies. As mentioned in chapters 1 and 2, online marketing, like any marketing process, operates effectively only when using accurate customer information. Companies that offer online marketing services collect huge amounts of user data under the guise of serving their customers better, and most importantly free of charge for their customers. Still many studies show that online privacy is really important for people and they would like to have a sense of control over which party is collecting information, and how it can be observed. However, online data collection differs from typical customer information collection by making possible to gather very detailed information about people, sometimes very invisibly, and the information might be hard to keep out of the hands of third parties.

The first recommendation is that the companies and service providers should provide people with better information about their data collecting methods. According to studies, people do not bother reading through the terms and privacy policies of the website, even though privacy-related values are very important to customers. In that case, companies should have a responsibility to provide information that is easier to read, and therefore easier to understand. One suggestion for sharing clearer information could be that before using any website - and even again from time to time - people would be forced to read its terms and privacy policy. Transparency and reliability go hand in hand. Customers want to have a sense of control over the collected data. Therefore it would be important that customers would have a chance to view the data collected from them, for what purposes data is used for, and even to have possibility to delete information companies have collected.

The last part of the survey focused on finding out how much people trust data practises of the companies. Results showed that people do not trust companies to keep customer information safe, and if the information leaks to other parties, people's confidence decreases. Based on the survey results and many studies, people really care about who is collecting information and who will use it. Therefore a recommendation for companies is to truly respect the privacy of people and make sure that the collected information does not end up in the wrong hands. Another suggestion for the companies is to only collect user data that is necessary for their operations, not too detailed

information. The companies should store user data on internal hard disks, and if the data is forced to be used externally, it has to be masked somehow so that no one can be identified.

Overall, people should be provided with better knowledge of privacy issues. Not just of online viruses, worms and spywares, but also letting them know how many popular websites have a third-party presence, and therefore how people could prevent and reduce the spread of their data. This should be the responsibility of not only the internet companies, but also legislators.

CONCLUSION

This bachelor´s thesis studies people´s knowledge of online data collection and whether or not it decreases trust in companies that collect data. The purpose of the research was to find out how aware people are that their online privacy might be in danger every time they are using the internet. To support the purpose of the study, the following research questions were formulated:

1) What is people's general knowledge about data collection?
2) How does online data collection affect people's trust in companies?
3) Do people trust in companies´ ability to maintain customer information safely?

Theoretical framework for this study consists of previously published articles, studies and literature in the field of data collection. Previous studies and articles showed that many popular websites have a third-party presence and many third-party cookies are owned by a handful of companies. They also showed that people have great worries about online privacy and therefore confidence towards to companies that are collecting user data is low. Previous studies also showed that people do not know how extensive the collecting is and that data is collected even when they are not on the company's website. Literature about online marketing and its aspects were introduced in order to understand how important online data collection is for online marketing, online advertising and targeted marketing.

As for the methodology of study, both quantitative and qualitative research methods were used to get an understanding of how common this data collection has become and how much people know about it, and how they feel about companies doing so. Theories from previously published articles and studies were used as a basis to formulate the research questions. The survey - which was created to gather data to answer the research questions - gave a good view of how much people know about online data collection and how it affects their trust towards companies.

The main findings of this research show that people are aware that companies and service providers are collecting user information and that they are afraid of their privacy if the data gets into the hands of third parties. People will not accept this and if the data gets to other parties, their trust decreases. Based on the survey results, participants were aware that companies collect detailed

information about them for marketing purposes. It implies that people have seen notifications of cookie usage on websites, and therefore signals that companies and service providers are providing information about the usage of cookies on their websites. Results also showed that people were not sure if websites´ privacy policies would prevent data from getting shared with other parties. People are most likely unaware of how different web cookies of same the origin share information with each other – commonly referred to as "cookie syncing". In the name of transparency, this information should be provided to people by companies and service providers. Survey results show that people feel that their privacy is in danger when companies collect information of them and their web habits. This is a sign that people do not trust how companies handle data. This fear may be appropriate, as many data leaks have occurred in the past. Therefore, companies should make sure that collected information does not fall into the hands of third parties. The survey results imply that people's trust in companies will fall if user information is leaked. Therefore, when the company is collecting user information, it must be treated with respect and care.

Limitations of this research are related to the sample size of the survey, which limits the generalizations that can be made on the basis of the results of this survey. Another limitation is that the survey might include non-representative participants.

For possible further research, survey questions could perhaps be extended to inquire more about how extensive data collection people accept. If seems that companies can serve people better with the assistance of cookies. The important question is where the limit between customers´ privacy and the fluidity of their cookie-assisted web surfing experience lies. Web companies must find a way to make the most of the use of cookies without making people feel vulnerable.

Overall, this study shows that people are aware of service providers collecting information and they understand how web cookies work for the data collection and online marketing. However, people don´t like the idea of their data getting into the hands of third parties, and the author´s study has shown that this worries people. Furthermore, findings showed that people's trust in companies' will decrease if collected user data is exposed to third parties. These results are pretty similar to previous studies.

As a recommendation, companies should provide a clear overview of what data they are collecting and for what reasons, and this should be done in a way that really forces the website user to understand data collection. This could prevent unpleasant surprises when the use of third-party

cookies is exposed to the user. People don´t trust companies to handle their data with care and this decreases confidence towards companies. In this case, companies should make a bigger effort to convince people to understand why information is gathered and how companies can protect the information from other parties. Customers want to have a sense of control over the collected data. Therefore it would be important that customers would have a chance to view the data collected from them, for what purposes data is used for, and even to have possibility to delete information companies have collected. Furthermore, survey findings show that people do not trust companies that have leaked user information. Further areas to research could be how much data collection people accept at the expense of their privacy. This would be important for companies so that they can better understand their boundaries in data collection. Governments should also stay on top of the situation so that customers can use the Internet without a fear of their privacy being compromised.

REFERENCES

Awad, E. M. (2006). Electronic commerce: from vision to fulfilment. 3rd ed. New Jersey: Pearson Education, Inc.

Charlesworth, A. (2007). Key concepts in e-commerce. New York: Palgrave Macmillan.

Chaffey, D. (2004). E-commerce and E-management: strategy, implementation, and practice. 2rd ed. Essex: Pearson Education Limited.

Kotler, P., Armstrong G., Harris, L. C., and Piercy, N. (2013). Principles of marketing. 6rd ed. Essex: Pearson Education Limited. P. 109

Linton, I Chron (31.1.2018) Six Benefits of Internet Marketing
http://smallbusiness.chron.com/six-benefits-internet-marketing-31382.html

Google website- 28.11.2017

Castelluccia, C and Narayanan, A (2012) Privacy considerations of online behavioural tracking p.4

Rosche, E (2016) How Does Google Analytics Collect Information?
https://www.lunametrics.com/blog/2016/06/22/google-analytics-collects-information/

MND web docs (2018) https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies
4.5.2018

Markelz, M (2016) Digital Tracking Technologies: A Primer
https://www.ama.org/publications/MarketingNews/Pages/digital-tracking-technology-basics.aspx

Engelhardt, S and Narayanan, A (2016) Online Tracking: A 1-million-site Measurement and Analysis
http://randomwalker.info/publications/OpenWPM_1_million_site_tracking_measurement.pdf

Finlex (2011) http://www.finlex.fi/fi/laki/kaannokset/2004/en20040516.pdf
Section 7

Humbries, D (2014) Public Attitudes Towards Data Collection and Privacy IndustryView | 2014
https://www.softwareadvice.com/security/industryview/data-privacy-report-2014/

Madden, M and Rainie, L (2015) Americans' Views About Data Collection and Security
http://www.pewinternet.org/2015/05/20/americans-views-about-data-collection-and-security/

Turow J., Mulligan, D. and Hoofnagle, C. (2007) Consumers fundamentally misunderstand the online advertising marketplace

Falahrastegar, M., Haddadi, H., Uhlig, S. and Mortier, R. (2014) Anatomy of the Third-Party Web Tracking Ecosystem
https://arxiv.org/pdf/1409.1066.pdf

Lewis, S.J. (2017) The Information Superhighway has become The Information-Tracking Superhighway
https://mascherari.press/the-information-superhighway-has-become-the-information-tracking-superhighway-2/

Wills, C. (2011) Most Major Websites Leak Private Data, Study Finds
https://www.businessnewsdaily.com/1040-web-sites-data-leaks-privacy.html

Kissmetrics (2014) https://blog.kissmetrics.com/behavioral-advertising/
(source from

Madden, M. (2014) Public Perceptions of Privacy and Security in the Post-Snowden Era
http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/

Directive on privacy and electronic communication (2002) Article 5(3)

Appendix 1. Online Survey

General knowledge about data collection and its trustworthiness

Aim of this survey

This survey was created to figure out what people think/feel about companies that gather detailed information about their customers. Huge amount of personal data can be collected from the visitor of firm's website, varying from location to personal interests. Companies usually use customer data to improve the user experience of their service and also to make their marketing efforts more efficient. This means that companies store big amounts of customer data which they have to keep safe, but sometimes personal data get into the wrong hands.

1. Gender

- Female
- Male

2. Age

3. Are you aware that every time you visit a company's website, the company may collect information about your movement's there and even detailed information about you (such as personal interests)?

- Yes, I'm aware
- No, I'm not aware
- I don't care

4. When browsing internet, have you ever paid attention to notifications of companies using cookies on their web site?

- Yes I have
- No I haven't

5. When a web site has a privacy policy, I know that the site will not share my information with other websites or companies.

- True
- False
- I don't know

6. Corporations exploit users' web habits by tracking cookies to collect information about buying habits. It's also possible to find out which pages the user has visited, in what sequence, and for how long.The data can then be collected and sold to bidding corporations. Do you accept this?

- Yes
- No
- I don't care

7. Do you feel that your privacy may be in danger when companies collect information about you and your web habits?

- Yes, it worries me
- No, it does not worry me
- I don't care

8. Are you afraid that customer data collected by corporations might be exposed to third parties?

- Yes I am
- No I'm not
- I don't care

9. Do you trust companies to handle customer information confidentially?

- Yes
- No
- I don't care

10. Would you trust a company that has accidentally leaked customer information to a third party?

- I don't trust at all
- My trust has decreased
- My trust has remained the same
- My trust has increased

Appendix 2. Quantitative Data

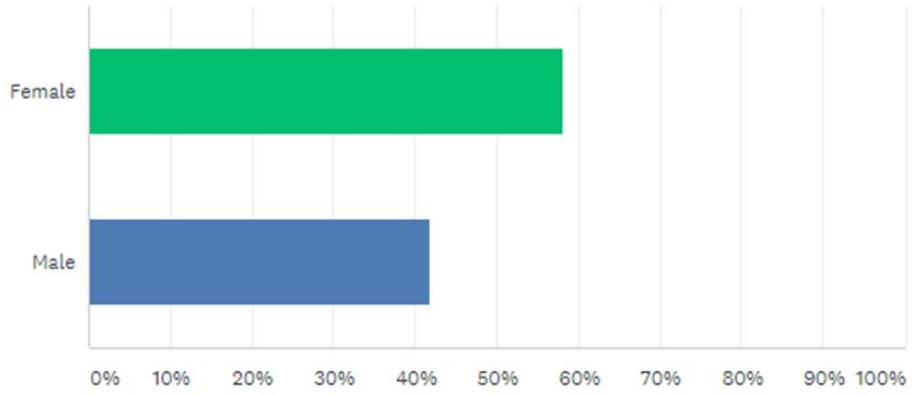

Table 2.1 Respondent's gender

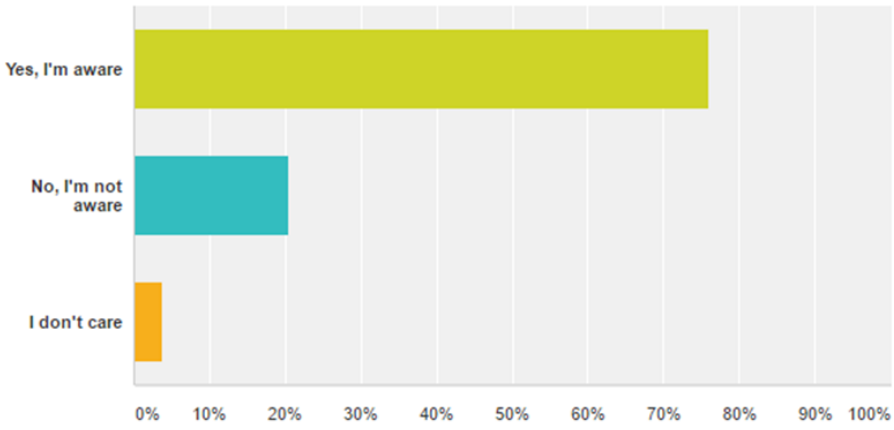

Table 2.2 General knowledge about data collection

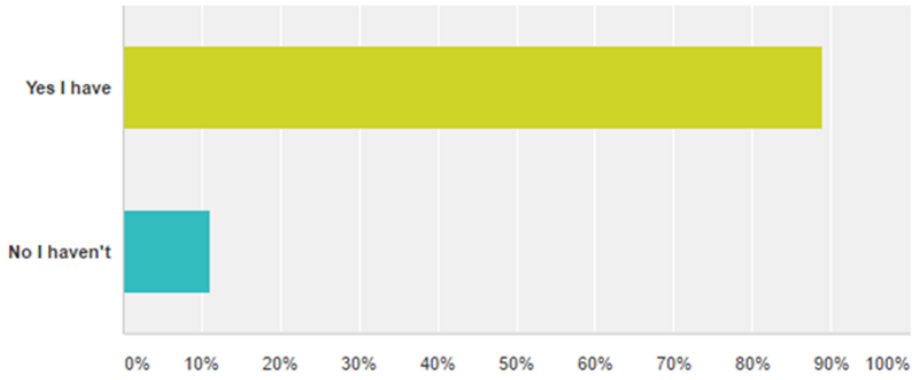

Table 2.3 Have you ever paid attention to notifications of firms using cookies on their website

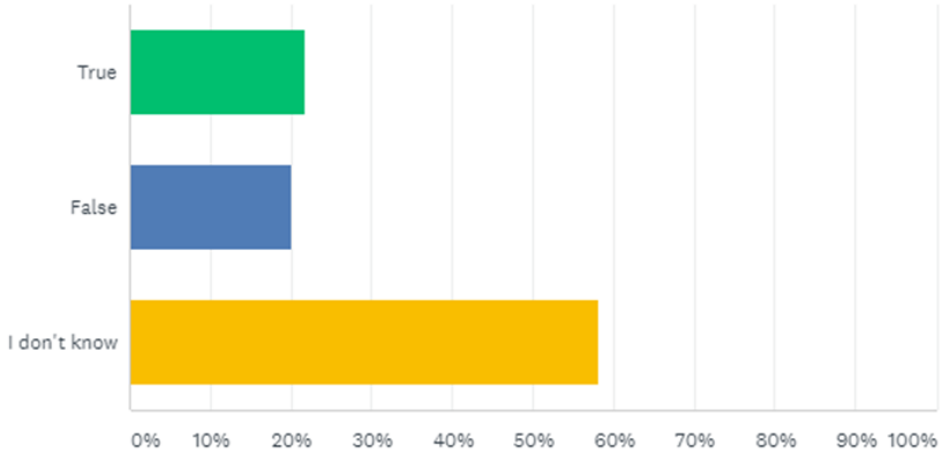

Table 2.4 When a web site has a privacy policy, I know that the site will not share my information with other websites or companies

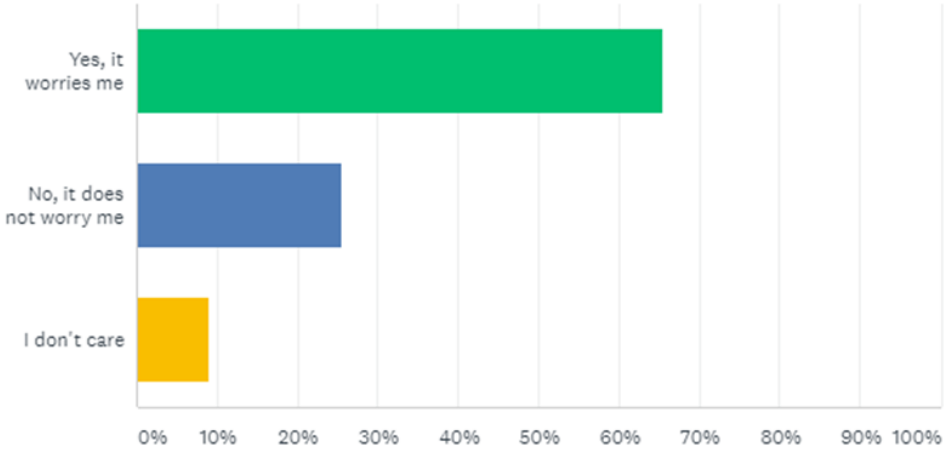

Table 2.5 Do you feel that your privacy may be in danger when companies collect information about you and your web habits

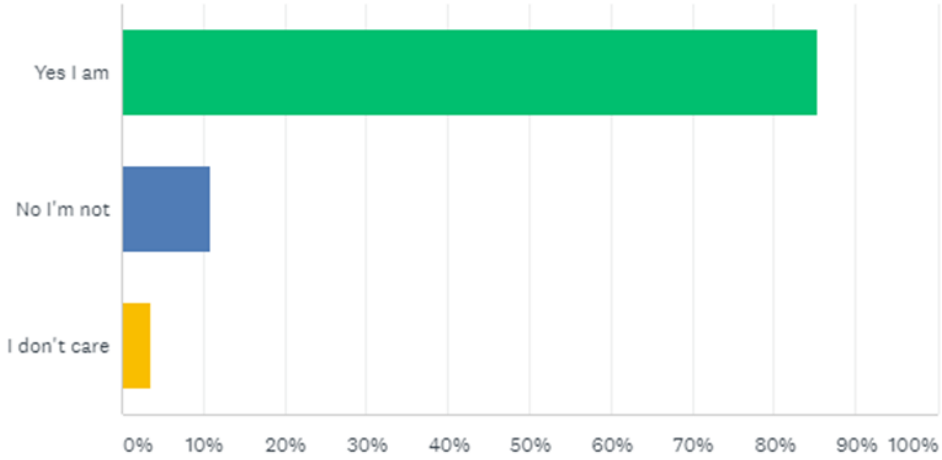

Table 2.6 Are you afraid that customer data collected by corporations might be exposed to third parties

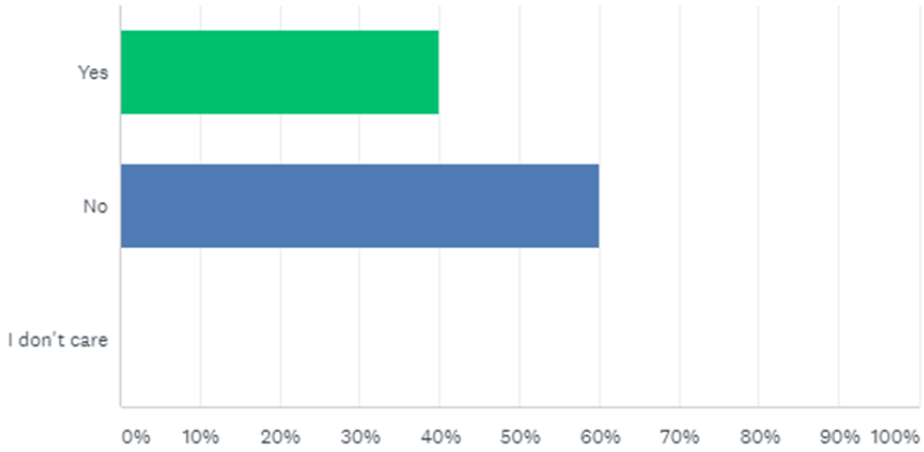

Table 2.7 Do you trust companies to handle customer information confidentially

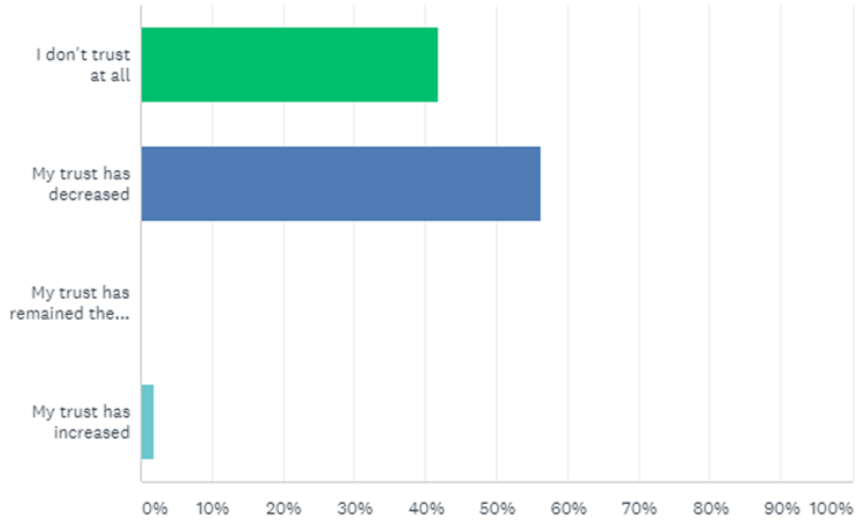

Table 2.8 Would you trust a company that has accidentally leaked customer information to a third party