Bachelor thesis
Computer Science

Radboud University

Privacy Protection Against Cross-Device Tracking Methods

Author:
Tom Nies
s4345746

First supervisor/assessor:
dr. Gergely Alpár
gergely@cs.ru.nl

Second assessor:
dr. Antonio de la Piedra
a.delapiedra@cs.ru.nl

June 19, 2016

Abstract

The amount of mobile devices is growing every day. People also tend to switch between their devices more frequently. Because of this, their data becomes fragmented across multiple devices. New tracking techniques are being developed to be able to track users across their devices. This thesis discusses three cross-device tracking methods and analyses their privacy aspects.

Contents

Chapter 1

Introduction

The amount of people that own a mobile device grows every year. In 2015 the total number of mobile subscriptions was 7.4 billion [10, 26]. It is expected that this number will grow even further, up to 9.1 billion mobile subscriptions in 2021 [10]. Because of this vast amount of devices and the ease of use, people tend to switch between devices frequently, and user data becomes fragmented across the different devices they use.

When someone is using only one device, all their data will be on this device. Since all the data is on one device, it will be easy to identify this user. However, people are now using more than one device, which makes it harder to link a user to all these devices. The data that is fragmented across multiple devices has to be combined to be able to identify a user. The technique used for this is cross-device tracking.

The goal of cross-device tracking is to determine whether different devices (for example a computer, smartphone and tablet) belong to the same user. This information is especially valuable for companies in the advertising industry, because the information can be used to show targeted advertisements to the user (on all their devices). Targeted advertising is known to be more effective than random advertisements [6] and can thus be used to increase profits.

However, cross-device tracking also raises privacy concerns. First of all, tracking might cause chilling effects: people change their behavior if they know they are being tracked. Second, users do not have control over their personal information. They do not know which information is collected, how

this information is used and with whom the information is shared. Third, there is a risk of discrimination or manipulation based on collected data [27]. In this thesis we will take a look at three different cross-device tracking methods and analyse their privacy aspects. The question that this thesis answers is: How can we achieve more privacy protection against cross-device tracking methods?

The rest of this thesis is organised as follows. In chapter 2, we will discuss the background of cross-device tracking. Chapter 3 discusses the three cross-device tracking techniques that we are going to analyse in chapter 4. In chapter 5, we discuss the implementation I made for the SilverPush technology. The related work is presented in chapter 6. In chapter 7, we present the conclusions of this research.

Chapter 2

Preliminaries

In this chapter we discuss what cross-device tracking is and how it can be used. We also take a look at the users' control over their data and user awareness.

2.1 Cross-device tracking methods

Cross-device tracking is used to find out which devices belong to the same user. Assume you own a laptop, a tablet and a smartphone and use them regularly. You decide that you want to buy a pair of shoes and use your tablet to search for new shoes online. When advertising engines observe your intention, normally this will cause you to see advertisements for shoes only on your tablet. However, an advertising company might find it interesting to also show these advertisements for shoes on your laptop or smartphone. To determine which laptop and smartphone are yours, cross-device tracking is used.

We can distinguish two different types of cross-device tracking methods: deterministic and probabilistic [7, 8].
Definition. Deterministic cross-device tracking is a method to connect multiple devices based on a persistent unique identifier.
Definition. Probabilistic cross-device tracking is a method to connect multiple devices based on inferences about likely connections between devices

or users.

The difference between these methods is that probabilistic methods are based on the chance that specific devices belong to the same user based on collected data. Deterministic methods on the other hand provide certainty that some devices belong together and are based on data provided by a single user.

2.2 User control

Another distinguishing factor is the control that a user has over each of these methods. This means that users have the possibility to choose what they share or not. This gives users the ability to change the behaviour of the method.

These two dimensions, the methods and the control, form the following matrix.

	Control	**No control**
Deterministic	1	2
Probabilistic	3	4

Figure 2.1: Method-control matrix

To clarify these four possibilities we discuss each of them and give some examples.

1. In this case deterministic cross-device tracking is used, and the user does have control. Because deterministic cross-device tracking is used, the user has to give away some identifying information about himself. However, users have control over the information and they can choose not to share the data.
   An example of this is when you own an Android tablet and an Android smartphone. When you use these devices for the first time, you will have to log in with your Google account. By doing this, you are telling Google that you own these devices. Google will then know that all the

data they collect on this tablet and smartphone belongs to you. This allows Google to use deterministic cross-device tracking. Yet, users have the ability to log out or not even log in in the first place, so they have control over the information collection.

2. In this case deterministic cross-device tracking is used, and users do not have control. This means that users have to give away some information about themself, but they do not consciously choose to do so.
   Imagine that there is a big company which owns dozens of websites. You use your e-mail address to subscribe to the newsletter of one of these websites on your laptop and also to the newsletter of another website on your smartphone. The company now sees that you have subscribed to both newsletters with the same e-mail address from multiple devices. If a cookie is set on both devices, the company will be able to identify your devices and connect these with the e-mail address you provided. You do not have the possibility to influence this, because you do not know that the websites belong to the same company.

3. In this case probabilistic cross-device tracking is used, and the user does have control. This means user data is collected to connect multiple devices based on inferences of the collected data. Moreover, the user has the ability to control the collected information.
   An example of this is when you are travelling and you carry your smartphone and laptop with you. If you are using navigation software on both devices and GPS is turned on, then the locations of both devices will correspond for a long period of time. Based on these corresponding locations it can be concluded that both devices belong to the same person. Since you have the possibility to turn GPS on/off, you are in control of the information collection.

4. In this case probabilistic cross-device tracking is used, and the user does not have control. This means that user data is collected to connect multiple devices, and the user does not have a possibility to restrain data collection.
   For example, a computer can emit a high pitch that is inaudible for humans. However, another device can receive this high pitch. If a computer emits a high pitch on a regular basis, it can be determined which devices are often near the computer. Based on this information

it can be concluded that a certain device belongs to the same person as the computer.This method is probabilistic, because it does not provide certainty as the devices near the computer can be different everytime. However, based on the collected data it is possible to determine with increasingly higher probability which devices belong to the same person. Users do not have control over the data collection and the collected data. Moreover, users are often not even aware that this technology is used to collect their data, because they do not know about this technology's existence and they can not hear the emitted sounds.
This technology is discussed in detail in section 3.1.

2.3 User awareness

Having control over your own data is important, but it is not enough. Users should also be aware of the fact that data is collected about them and they should consider whether or not they want to share certain information for particular purposes. An example of user awareness is logging in to Facebook. By logging in you identify yourself and you are aware that you might give away personal information to Facebook. On the contrary, a user might not be aware of giving away information if a company collects small pieces of personal information to profile a user. For example, if an advertising company knows that you regularly visit websites about cars and football, then they might conclude that you are a man and show you advertisements based on this aggregated information. Users might not be aware of this, because they often do not know from which websites the information is gathered.

Chapter 3

Cross-device tracking methods

In this chapter we discuss three cross-device tracking methods that are used in practice. The first technique uses inaudible audio beacons to send ultrasonic messages across devices. The second method is a platform that enables companies to work together to identify their customers. The third technique, Kraken, uses software to collect as many information as possible from all devices owned by the user.

3.1 Inaudible audio beacons

SilverPush is an Indian company located in New Delhi and San Francisco. SilverPush develops software to track users across multiple devices. Its goal is to deliver accurate matching based on audio bridging [24]. SilverPush provides a software development kit (SDK). Any application that includes this SDK can be used to track users.
The users can be tracked across multiple devices through the use of inaudible audio beacons that emit ultrasonic sound. Audio beacons are triggered when you see an advertisement. So if you see an advertisement on your laptop, the advertiser will place a cookie on your laptop and simultaneously play an ultrasonic audio sound by using the speakers of the laptop. Another device, your smartphone for example, will be able to hear and recognise this

sound if the right software is installed. The software will then place a cookie on your smartphone with the same cookie value. Because these cookie values are equal, it will link the two devices [18].

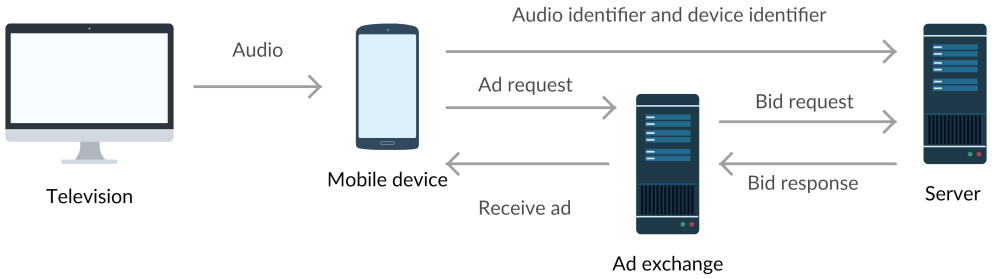

Figure 3.1: SilverPush's infrastructure [15]

In figure 3.1 we can see in detail how this technology can be used in practice. If you are watching television and there is a commercial break, then some of the advertisements can include ultrasonic messages. The audio of the television can be received by one or more other devices by the SilverPush SDK. This is depicted in figure 3.1 as one mobile device, but there can be multiple devices. The received audio contains an audio identifier. This audio identifier is a unique identifier for a certain television advertisement. The mobile devices filter the identifier from the received sound and store it. If the same identifier is stored on multiple devices, the devices can be linked. The audio identifier and an ID of your device are send over the internet to the SilverPush server. The device identifier can be a device ID, IMEI number, Android ID, Apple ID, IDFA (Advertising Identifier), AID (Advertising Identifier), UDID (Unique Device Identifier), mobile number or MAC ID [15]. The audio identifier and device ID are stored in a database when the server receives them.

If you use your mobile device at a later moment and use an application that includes an advertisement slot, then your device sends an ad request to an ad exchange. The ad exchange then sends a bid request to various servers. The SilverPush server will then check if there is an entry in the database for the device ID in the bid request. After finding the corresponding database entry, the server will send a bid response to the ad exchange. In the last step, the mobile device receives an ad and shows this ad to you. The ad will

be related to the television advertisement.

The ultrasonic messages sent in the first step vary in range between 18kHz to 19.95kHz [25]. Humans are not able to hear these sounds, but our devices can. Kevin Finisterre of security consultancy Digital Munition discovered that the high-pitch tones translate into characters. For example: an 18kHz sound is translated to an 'A' and an 18.3kHz sound is translated to an 'E' [12]. The sending side translates each character of the data into frequencies and emits sounds at those frequencies. The receiving side has to do a continuous Fourier transformation to find peaks in the received signal [15]. After finding a peak, the receiver translates the frequency back to a character.
A television advertisement has a two-letter sequence as an identifier [12]. When this technology becomes widely used, this sequence will probably be longer, because there are not enough two letter combinations to give every advertisement a unique identifier.

Addons Detector published a list of all the applications that use SilverPush's audio beacon system or used it in the past [9]. Most of these applications are run by Indian companies.
In chapter 5 we discuss the implementation I made for the SilverPush technology.

3.2 Adobe Marketing Cloud Device Co-op

Adobe recently started working on their own cross-device tracking platform. Adobe's cross-device co-op allows companies to work together in offering personalised services to their customers [3]. Companies all share there knowledge, so everyone can benefit from the available data [20, 5]. Adobe's role is to offer a platform that makes it possible for companies to share their data.

The most accurate method link someone's devices is by using login information [14]. For example, you log in to the website of co-op member A on your laptop and your smartphone and to the website of co-op member B only on your laptop. A is able to link your laptop and smartphone, because you

used both of these to log in to A's website. So A knows that this laptop and smartphone belong to you. At a later moment you use your smartphone to log in to the website of co-op member B. B only knows your laptop, not your smartphone. However, A knows both the laptop and smartphone and shares this information, via Adobe, with B. Co-op member B is then knows the smartphone should be associated with your laptop and can adapt to this. In exchange for sharing this data, A also receives similar data from B and the other co-op members. So sharing the data can be beneficial.

The co-op members give Adobe access to hashed login IDs and HTTP header data [3, 20]. Adobe processes this data and creates devices clusters (groups of devices that are linked together). It is unknown how this works exactly. Contacting Adobe about the details of their technology was a futile attempt. When a co-op member sees a new device, Adobe tells the member about the device cluster. So the co-op member only have access to the information that is relevant for them. Moreover, no personally identifiable information (information that can be used to identify a single person, like a name or passport number) is being shared about individual users [20, 19]. Only information about devices is shared and the relation between those devices.

3.3 Kraken

Kraken is a multi-device user tracking suite developed as a research project by the German TU Darmstadt [22]. The goal of this research is to get a better understanding of human activity. The results can be used by companies to offer personalised services. Examples of these services are personalised advertisements and personal assistants like Microsoft's Cortana or Apple's Siri. A lot of information is needed to improve this understanding of human activity. The Kraken framework offers software to gather this information from multiple devices. Users might want to allow such extensive data collection, because companies can use the data to offer them more accurate personalised services.
The Kraken framework consists of multiple components. As can be seen in figure 3.2, the central component is the Kraken server. The server's main activity is to store all the collected data. Besides that, the server also analyses all the data. The actual data collection happens in three places. First

of all, there is the desktop monitor which observes a user's interaction with his Windows computer. Secondly, there is the mobile monitor which studies how people use their smartphones and tablets. Thirdly, users can connect their social media accounts to Kraken. This gives Kraken the possibility to gather data from these social media accounts. The last component of the Kraken framework is the Kraken webpage. This webpage allows users to download the previously mentioned software.

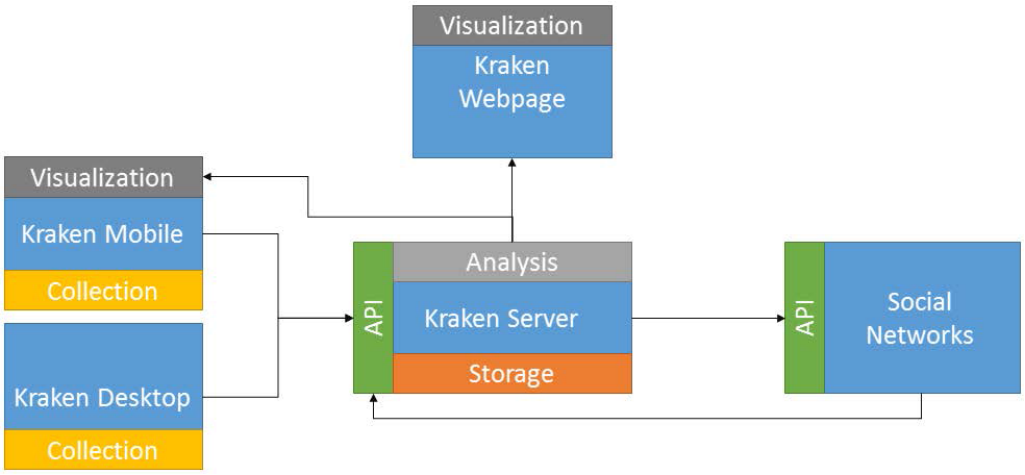

Figure 3.2: Kraken's architecture [22]

3.3.1 Kraken server

The Kraken server's main task is to store all the collected data. The data is stored in flat tables to keep the raw data in its original structure [22]. The raw data is not anonymised, because this is not possible given the vast amount of personal data. The data in these tables can later be analysed and the results can be visualised in graphs.

3.3.2 Desktop monitor

The Kraken Desktop Monitor observes the interaction of users with their Windows computer. It captures the processes running in the foreground and how a user interacts with these processes (e.g. mouse clicks in a user

interface). For some programs even more data can be captured. For the most common file types, like .pdf, .docx, .xlsx and .pptx, the Desktop Monitor can even extract data while people are working on them. Other data collection methods involve reading the clipboard content or checking the websites users view in their browser.

3.3.3 Mobile Monitor

The Kraken Mobile Monitor consists of an Android app that monitors a user's interaction with his smartphone. A smarthphone has a lot of physical sensors that can be used to collect data. It has for example access to an accelerometer, light sensor, proximity meter and location [21]. Besides the physical sensors, there is also a lot of software with useful data (e.g. e-mail, calendar). This makes a smartphone perfect for data collection.

3.3.4 Social networks

Kraken can also collect data from social media if a user connects his Kraken account with his social media account. Kraken will then use the social media's API to crawl each account once a week. This allows Kraken to collect profile data, location data, lists of friends, photos, messages, etc.

3.3.5 Kraken webpage

The webpage, kraken.me, allows users to download the desktop and mobile monitors. Users can also use the webpage to connect to their social media accounts. This is planned to be extended to offering personal assistance and providing the user with a rich visualisation of interaction data [22].

Chapter 4

Privacy analysis

In this chapter we discuss the privacy aspects of each of the cross-device tracking methods described in chapter 3. To analyse these methods we use Solove's taxonomy [23].

4.1 Solove's taxonomy

Daniel J. Solove [23] proposes a taxonomy of privacy, a framework that attempts to systematically separate and identify privacy problems. The taxonomy is split into four categories: information collection, information processing, information dissemination and invasion. In the next sections we briefly discuss the first three categories. The last category, invasion, will not be covered in detail here, since we are mainly interested in how the information is collected and what happens with this data. Invasion does not necessarily involve information, so it is less interesting for us. So instead of discussing invasion in detail, we take a look at how parts of the other categories can be exploited.
As you can see in figure 4.1, there is a data subject. The data subject is the person whose information is collected. After collecting the information, the data is in possession of the data holders. The data holders can process the collected information and can distribute the information to other people or organisations. The arrows show the flow of information and the data subjects' control over his information. If the data moves further away from

the data subject, the data subject loses increasingly more control over the information. In the next part we shortly explain the categories that are most relevant for the analysis.

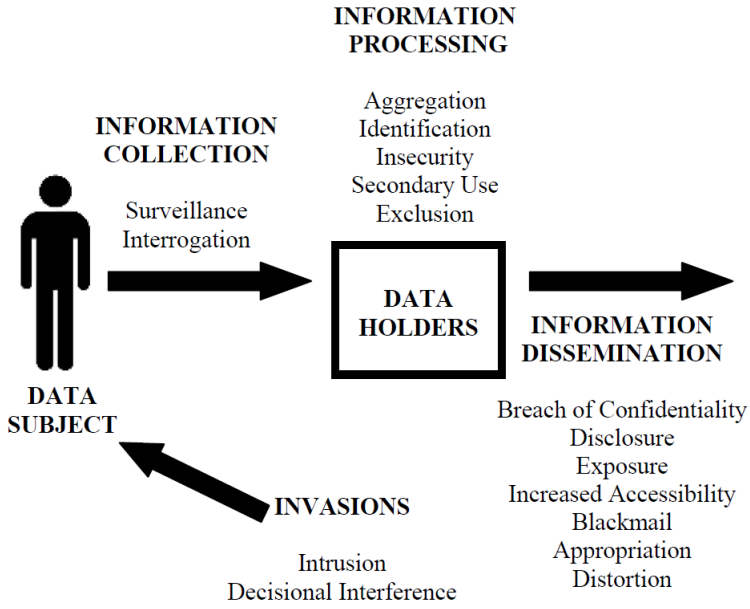

Figure 4.1: Solove's taxonomy [23]

4.1.1 Information collection

The first category is about the collection of information. Information can be collected in two ways: surveillance and interrogation. Interrogation does not apply to the cross-device tracking methods we've seen, because there is no questioning or probing of the user involved in using this technology.

- **Surveillance** is the monitoring of people's activities and behaviour.

4.1.2 Information processing

Information processing consists of actions that process the collected data. The data can for example be stored or manipulated.

- **Aggregation** is the combination of various pieces of information which leads to new information.

- **Identification** is the process of linking information to a person.
- **Insecurity** is the irresponsible handling of user information which makes it vulnerable to leaks and unwanted access.
- **Secondary use** is the use of information for something else than it's original purpose.
- **Exclusion** happens when the data subject does not know what data others have about him and doesn't have access to this data.

4.1.3 Information dissemination

Information dissemination involves the transfer and spreading of information.

- **Breach of confidentiality** is violating someone's trust that you do not disseminate his personal information.
- **Disclosure** is the spreading of truthful information about someone. This can cause reputational damage.
- **Increased accessibility** is increasing the accessibility to someone's personal information.
- **Blackmail** is the threat to disclose someone's personal information.
- **Distortion** is the spreading of false information about someone.

4.2 Inaudible audio beacons

4.2.1 Information collection

The goal of SilverPush's technology basically is to collect information. The inaudible audio beacons can be used to determine which devices are in close proximity to each other. If two or more devices are often near each other, they probably belong to the same person and these devices can be linked. On the other hand, the inaudible audio beacons can also be used to collect information about what someone is watching on television. Or more generally, what they are looking at on their screen.

The SilverPush technology is now used to determine which advertisements people see on the television, but this technique can also be used in other ways. Instead of a television emitting the ultrasonic sounds, this can also be done by a laptop speaker or smartphone. One possibility to use this is to include a unique audio beacon into every page of a website. By doing this, a user can be tracked across the internet and is under constant surveillance. Including audio beacons into webpages can be extended to Tor [13], which might help de-anonymising Tor users.
Another possibility is to include a unique audio beacon in an illegal video. When someone watches this video, the audio beacon will be triggered and the ultrasonic sound can be received by a smartphone. This smartphone could then send a message to the police reporting the illegal video is watched. Additionally, by including the IP address or GPS location in the message, the person watching the video can be identified.

4.2.2 Information processing

The information collected by SilverPush's technology can be aggregated to gain extra information. The audio beacons can be used to link multiple devices, but can also be used to determine which advertisements have been watched. This information can be combined to show additional advertisements on all of the linked devices.
Identification based on the audio beacon technology as it is used now is not possible. The technique is used to link device and no personally identifiable information is used to do this. However, identification can be possible if the audio beacons are deployed in ways that do involve personally identifiable information. If the audio beacons are used on the internet, as suggested earlier, then they can be used to gather the needed personal information. For instance when someone includes a unique ultrasonic sound message in a video sent by e-mail. If the video is played, the ultrasonic sound is received by a smartphone. This smartphone can then send a message back to the sender of the e-mail and the sender will know that the e-mail address belongs to the same person as the smartphone. This way the owner of the devices can be identified. Another way to include personally identifiable information is to include the phone number or other identifier in the message that the smartphone sends to the server (figure 3.1). A phone number is unique

for a person and can thus be used to identify someone.
Insecurity, the carelessness in protecting stored information from leaks and improper access [23], often leads to identity theft. This is not possible if only information about the linked devices is leaked, because this does not contain any personally identifiable information. However, if the information leaks about what someone watches on the television every day, then this information can be used in combination with other information to build a profile about this person. If this profile is detailed enough, it can provide enough information to enable identity theft.
Most people are not aware of this technology's existence. And even if they were, they would not notice when it is used, since they are unable to hear the ultrasonic sounds. The software needed to receive and recognise the ultrasonic audio can be hidden in a legitimate application [24, 9]. This leaves the users without any knowledge that they are being tracked, due to a lack of transparency. They are completely excluded from the process. Besides this, users do not have access to the collected data and are not able to make corrections to any inaccurate data. Because of this, there can be false positives in the linking of devices. An example of this is when your friend comes over to watch television at your place. Your friend's smartphone will receive the same ultrasonic sounds emitted by the audio beacons. This might cause his smartphone to get linked to your smartphone and he might receive advertisements that were actually meant for you. This can cause awkward situations if the advertisements are about sensitive subjects, such as your medical situation.
To improve the transparency of this technology a few easy adjustments can be made. It is for example possible to show the SilverPush logo in an advertisement when inaudible audio beacons are used. Another possibility is that your smartphones makes a sound, vibrates or shows a message that it received an ultrasonic sound. As a result of this, users become aware of the usage of this technology. The next step is to give users control over their own data. According to the American Center for Democracy & Technology (CDT) audio beacons are impossible for users to control [18]. This is not entirely true. Users can for example mute the television when there is a commercial break or they can turn off their internet so the smartphone is not able to send a message to the servers. Another and better possibility is to explicitly ask a user for confirmation before sending a message to the

servers. This way a user has control over his own data. It is possible to extend the user's control even further by giving users access to their own data stored on the server and giving them the opportunity to correct inaccurate data or delete (parts of) their own data.

4.2.3 Information dissemination

The data is moving further away from the data subject's control in the information dissemination part of Solove's taxonomy. However, in the case of the inaudible audio beacons, the user already has almost no control over his own data during the information collection and processing. Information dissemination can cause further privacy harms, for example the disclosure of what you are watching on television. However, this is out of reach for the users, so we will not discuss the details here.

4.2.4 Discussion

Users are often not aware that this technology is used, since they are not able to hear the emitted sound and there is no visual notification. Besides that, users have no control over their personal data. They do not know which information is collected, when this data is collected, what happens with the collected data and with whom the data is shared. When users become aware of this technology and they have no control over their data, chilling effects can occur. They might stop watching advertisements on television or turn off the sound during the commercial break.
Users should be notified when the inaudible audio beacons are used and they should have access to the collected data. SilverPush should offer users a way to view their data, edit their data and delete their data.

4.3 Adobe Marketing Cloud Device Co-op

4.3.1 Information collection

The information collection of Adobe's cross-device co-op starts when a user logs in to the service of one of the co-op members. The co-op member then shares the hashed login ID and HTTP-header data of this user with Adobe [3]. Adobe processes this data and shares the device cluster with the rest of the co-op members. The other members then know that someone used the service of another co-op member.
Adobe claims that users only participate in the cross-device co-op if they opt-in [14]. However, users automatically opt-in when they accept the privacy policy of one of the participating companies [20]. Most users accept such policies without reading them. As a result of this, users will not be aware that data about their devices is shared with other companies. Adobe does offer users a way to opt-out from participating in the cross-device co-op. A cookie is used to remember the user's choice to opt-out [20]. So if a user chooses to opt-out and then deletes his cookies, the opt-out is forgotten and the user is being tracked again.

4.3.2 Information processing

The goal of the cross-device co-op is to link a user's devices. Every company that participates in the co-op shares its own knowledge about the logged-in users. This data can then be aggregated to link a user's devices. Users might not be aware of companies exchanging their data, because they automatically opt-in [20].
This technique can also be used for other purposes, like personalised services or advertisements. An example of this is when co-op member A runs a webshop that sells shoes and co-op member B owns a news website. If you log in to the webshop of co-op member A, they notify Adobe that you logged-in. Adobe then shares this information with the other co-op members, including co-op member B. If you visit the news website at a later moment, they know you are interested in shoes and they can show you advertisements of shoes on their website.
Adobe only focuses on linking devices and building a device cluster. There

is no personal information needed for this process and Adobe claims that they do not collect any personal information [3, 19]. Without personal information it is not possible to identify someone, so identification of users is not possible with this technique.
Another claim Adobe makes is that they will provide users a way to see which devices are linked to the device they are currently using [3, 20]. Unfortunately they do not mention anything about the possibility to change incorrect data. This might mean that this option does not exist at all and you are not able to do anything if the device of another user is linked to your devices.

4.3.3 Information dissemination

The information Adobe receives from co-op members allows Adobe to build device clusters. Because Adobe knows from which co-op member they received the data, they can also make a good guess what someone was doing. However, Adobe only knows about the devices and does not have any personally identifiable information, so they can not link this data to a person. Information dissemination can therefore not result in reputational damage, blackmail or other harmful activities.

4.3.4 Discussion

Users automatically participate in the cross-device co-op if they use a service of one of the co-op members. There is a possibility to opt-out, but unfortunately the opt-out choice is saved in a cookie and is lost at the moment users delete their cookies. It would be better to have an explicit opt-in, rather than an automatic opt-in by default and the possibility to opt-out.
Users should always have control over their own data. This includes having the option to edit incorrect data. Adobe doesn't mention anything about the existence of this option, so it is unknown if Adobe offers this option. If this option does not exist, then it is not possible for users to edit an incorrect device cluster.
Another privacy concern is the one of the disclosure of a user's data. Adobe should guarantee that the information they receive is safe and is not shared with parties outside the cross-device co-op.

4.4 Kraken

4.4.1 Information collection

Kraken collects as much data as possible from every device that has Kraken's software installed on it. So a lot of personally identifiable information is collected by Kraken. This is not done surreptitiously, because users have to download and install the software on their devices before Kraken can start collecting information. Users consciously choose to install the Kraken software, so they choose to be monitored. Since users are actually giving away their own information, this is not the same kind of surveillance as we've seen in the analysis of the other cross-device tracking methods.

4.4.2 Information processing

The three Kraken monitors allow Kraken to collect a lot of personal information about users. This does not only involve personally identifiable information like a name and e-mail address. The Kraken Desktop Monitor and Mobile Monitor can also collect data about the websites users are visiting, the documents they work on and the mobile applications they use. By aggregating this enormous amount of data Kraken can build accurate profiles of their users. This helps Kraken to get a better understanding of the users.
Besides getting a better understanding of users, Kraken can also use the profiles to offer personalised advertisements or other personalised services.
Identification of users is very easy for Kraken. If users connect their social media accounts to Kraken, then Kraken immediately knows the users' names. Even if Kraken does not get access to the social media accounts, Kraken can easily identify users. The Desktop Monitor allows Kraken to read the contents of pdf and Word files. People often include their name in these documents, so Kraken can extract these and identify the users.
The Kraken server stores all collected data in its raw form, because they want to retain the data in its original structure [22]. It is unknown if the data is stored as plain text or encrypted, but if the data is not encrypted this might cause some serious privacy issues when the data is stolen. The Kraken server contains a lot of information about who someone is and what

they do every day. The stolen data can therefore result in identity theft of everyone using the Kraken software.
One of Kraken's design principles is openness [22]. Users have access to their own data and they are offered the possibility to delete their data. Another design principle is self-determination. Users decide which data Kraken collects about them. The Kraken software can also be set into a deaf mode which disables the data collection. In summary, Kraken offers extensive user control.

4.4.3 Information dissemination

Kraken collects and stores a huge amount of data. Since Kraken has the possibility to collect almost all data from all of a user's devices, Kraken basically knows everything about your digital life. People use computers and mobile devices so much that the stored data is a valuable source of information. This information can consist of the user's personal information, information about his personal live or work-related data.
This data also involves information that people do not want to become public knowledge. The dissemination of the data might cause some troublesome situations. For example, the spreading of personal data can result in identity theft. The dissemination of sensitive information about someone's personal life can lead to blackmail or attacks upon their reputation and consecutively change other people's judgment about this person. The dissemination of work-related data does not only involve the privacy of the person using Kraken, but also their colleagues and customers.

4.4.4 Discussion

Users choose to download the Kraken software and are aware of the fact that all their data is collected and processed by Kraken. In return, Kraken offers a lot of transparency and control to the users. Users can access their own data and have the possibility to delete (parts of) their data. Besides that, users can set the Kraken monitors in deaf mode to disable the data collection.
The main privacy concerns for Kraken users are the insecurity and dissemination of their data. The data collected by the Kraken monitors should be

transferred securely to the Kraken Server. Obviously the data on the Kraken Server should also be secure and can be encrypted before it is stored. In addition to secure transmission and storage, Kraken should also make sure that no user data is disseminated without the user's consent.

Chapter 5

SilverPush Demo

There is an app, *SilverPushDemoApp*, available for download in Google Play [1]. This app shows how the inaudible audio beacon technology is working. The app can receive ultrasonic sounds and shows an advertisement when the corresponding identifier is sent to the app. However, I found that there was no program that can generate the ultrasonic sounds, so I wrote a python program which can generate audio at the right frequencies.

In appendix A you can find the code I wrote to send the ultrasound to the SilverPush Demo App. The program takes a string as input, converts the characters of the string to frequencies and emits sounds at these frequencies. The SilverPush Demo App can receive these sounds and show the corresponding advertisements. By brute-forcing the possible identifiers, I found that the demo app shows an advertisement if one of the following strings is sent: ADA, AGA, AJA, AKA, APA, AUA. The demo works over a maximum distance of around 6 meters.

I also wrote a python program, which uses obfuscation, to disturb the working of the code from appendix A. The program emits random sounds in the same frequency band, but more frequently than the code that is used to communicate with the app. Because of this, the app won't be able to receive the right data and it stops working correctly. The code to disturb the SilverPush technology is available in appendix B.

Chapter 6

Related Work

Tracking users online can be done in multiple ways. Most of these techniques involve the use of cookies [2, 17]. Cookies are small pieces of data that can be stored on a user's computer. They allow websites to remember who you are. Evercookies even try to circumvent a user's attempt to delete cookies and try to respawn the cookies everytime they are deleted [16]. Since 11 June 2015 the Dutch cookie law (Artikel 11.7a Telecommunicatiewet) has been operative, because of the privacy issues with tracking cookies. The cookie law requires websites to ask their visitors if they accept third party cookies.

Cookies are stored on a user's computer, so they are limited to one device and they are browser-dependent. This can lead to inaccurate or incomplete user profiles [11, 4]. Cross-device tracking is not limited to one device and allows companies to track users across multiple devices. Cross-device tracking is therefore a more valuable method, because user information can be collected from multiple sources.

The collected data is often used to personalised services and personalised advertisements. It is beneficial for companies to show personalised advertisements to their customers, because it increases the profit compared to random advertisements [6]. Another application of personalised services is search engines utilising the data to personalise search results [11].

Chapter 7

Conclusions

In this thesis we looked at the privacy aspects of cross-device tracking. Cross-device tracking methods are used to track users across multiple devices. The tracking can be done in several ways. In this thesis we discussed three cross-device tracking methods: SilverPush's inaudible audio beacons, Adobe's Marketing Cloud Device Co-op and Kraken's Multi Device User Tracking Suite. After explaining how these methods work we analysed the privacy threats they pose using Solove's taxonomy.
The research question of this thesis is: How can we achieve more privacy protection against cross-device tracking methods? The main problem with cross-device tracking seems to be that people lose control over their own data and that they are often not aware that they are being tracked. In the case of the inaudible audio beacons, users do not know that they are being tracked and they do not have the possibility to access their data. Adobe does offer users a way to view their own data, but they do not mention anything about the option to edit incorrect data. Besides that, users automatically opt-in for Adobe's cross-device co-op when they use a service of one of the co-op members and the opt-out is stored in a cookie which is deleted at the moment they delete their cookies. The main privacy issue for Kraken is the insecurity or dissemination of user data, because Kraken collects and stores enormous amounts of user data. Users should be made aware of the privacy threats involved in cross-device tracking and should be offered more control over their data.

7.1 Future work

Cross-device tracking is a growing business and there are a lot more possibilities to track users across devices. Companies like Tapad and Drawbridge also offer cross-device tracking technologies. The privacy aspects of these methods can also be analysed.
In section 4.2.1 I suggested to use inaudible audio beacons to track people across the internet by including the beacons into web pages. Another option with inaudible audio beacons is to de-anonymise Tor users by including the audio beacons into Tor content. Future research could focus on protecting the user's privacy against such methods.

Bibliography

[1] Silverpush beacon demo app. https://play.google.com/store/search?q=silveredge&c=apps.

[2] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 674–689. ACM, 2014.

[3] Adobe. Adobe announces cross-device co-op to enable people-based marketing, March 22 2016. http://www.adobe.com/news-room/pressreleases/201603/032216AdobeCrossDeviceMarketing.html.

[4] Hassan Aljifri and Diego Sanchez Navarro. Search engines and privacy. *Computers & Security*, 23(5):379–388, 2004.

[5] Tim Anderson. Adobe will track you across all your devices with new co-op project, March 22 2016. http://www.theregister.co.uk/2016/03/22/adobe_will_track_users_across_devices_with_new_coop_project/.

[6] Howard Beales. The value of behavioral targeting. *Network Advertising Initiative*, 2010.

[7] Ricardo Bilton. Cross-device tracking, explained, 2016. http://digiday.com/publishers/deterministic-vs-probabilistic-cross-device-tracking-explained-normals/.

[8] Federal Trade Commission. Ftc cross-device tracking workshop, November 16 2015.

[9] Addons Detector. Silverpush android apps, November 19 2015. https://public.addonsdetector.com/silverpush-android-apps/.

[10] Ericsson. Erisson mobility report: On the pulse of the networked society, 2015. http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf.

[11] Anisha TJ Fernando, Jia Tina Du, and Helen Ashman. Personalisation of web search: Exploring search query parameters and user information privacy implications-the case of google. In *PIR@ SIGIR*, pages 31–36, 2014.

[12] Kevin Finisterre. Silverpushunmasked. https://github.com/MAVProxyUser/SilverPushUnmasked/commit/bc1dde934c0be02cfce72b7ea68d4a147ddd308d.

[13] Thomas Fox-Brewster. Meet the 'ultrasonic' tracking company privacy activists are terrified of, November 16 2015. http://www.forbes.com/sites/thomasbrewster/2015/11/16/silverpush-ultrasonic-tracking.

[14] Anthony Ha. Adobe announces new data-sharing effort for cross-device ad targeting, March 22 2016. http://techcrunch.com/2016/03/22/adobe-cross-device/.

[15] CA (US) Hitesh Chawla, San Francisco. Method and system for cross-device targeting of users, July 30 2015. US Patent App. 14/606/227.

[16] Samy Kamkar. Evercookie, 2016. https://github.com/samyk/evercookie.

[17] Jonathan R Mayer and John C Mitchell. Third-party web tracking: Policy and technology. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 413–427. IEEE, 2012.

[18] Katie McInnis. Cross-device tracking requires strong privacy and security standards, 2015. https://cdt.org/blog/cross-device-tracking-requires-strong-privacy-and-security-standards/.

[19] MeMe Jacobs Rasmussen. Privacy by design exemplified - adobe announces the adobe marketing cloud device co-op, March 22

2016. https://blogs.adobe.com/conversations/2016/03/privacy-by-design.html.

[20] Zach Rodgers. Adobe pitches marketers on a cross-device data co-op, but privacy is a snag, July 28 2015. http://adexchanger.com/online-advertising/adobe-pitches-marketers-on-a-cross-device-data-coop-but-privacy-is-a-snag/.

[21] I. Schweizer, R. Bärtl, B. Schmidt, F. Kaup, and M. Mühlhäuser. Kraken.me mobile: The energy footprint of mobile tracking. In *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on*, pages 82–89, Nov 2014.

[22] Immanuel Schweizer and Benedikt Schmidt. Kraken.me: Multi-device user tracking suite. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 853–862, New York, NY, USA, 2014. ACM.

[23] Daniel J Solove. A taxonomy of privacy. *University of Pennsylvania law review*, pages 477–564, 2006.

[24] SteamFeed. Silverpush launches cross-device ad targeting with unique audio beacon technology, June 9 2015. http://www.steamfeed.com/silverpush-launches-cross-device-ad-targeting-with-unique-audio-beacon-technology/.

[25] Iain Thomson. How tv ads silently ping commands to phones: Sneaky silverpush code reverse-engineered, November 20 2015. http://www.theregister.co.uk/2015/11/20/silverpush_soundwave_ad_tracker/.

[26] International Telecommunication Union. Ict facts & figures: The world in 2015, May 2015. https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf.

[27] Frederik Johannes Zuiderveen Borgesius. *Improving Privacy Protection in the area of Behavioural Targeting*. PhD thesis, University of Amsterdam, 2014. http://dare.uva.nl/document/2/154442.

Appendix A

SilverPush code

```
import winsound
import sys

duration = 1000 #time in ms

#frequencies in hertz
frequencies = {
        'a': 18000,
        'b': 18075,
        'c': 18150,
        'd': 18225,
        'e': 18300,
        'f': 18375,
        'g': 18450,
        'h': 18525,
        'i': 18600,
        'j': 18675,
        'k': 18750,
        'l': 18825,
        'm': 18900,
        'n': 18975,
        'o': 19050,
        'p': 19125,
        'q': 19200,
        'r': 19275,
        's': 19350,
        't': 19425,
        'u': 19500,
        'v': 19575,
```

```
        'w': 19650,
        'x': 19725,
        'y': 19800,
        'z': 19875
}

def send(freq, duration):
        winsound.Beep(freq, duration)

if __name__ == "__main__":
        input = ""

        if len(sys.argv) < 2:
                print("Use: python send.py data")
                sys.exit(0)
        else:
                input = (sys.argv[1]).lower()

        i = 0
        while i < len(input):
                send(frequencies[input[i]], duration)
                i += 1
```

Appendix B

Disturb SilverPush

```
import winsound
import random

duration = 100 #time in ms

#frequencies in hertz
frequencies = {
        '1': 18000,
        '2': 18075,
        '3': 18150,
        '4': 18225,
        '5': 18300,
        '6': 18375,
        '7': 18450,
        '8': 18525,
        '9': 18600,
        '10': 18675,
        '11': 18750,
        '12': 18825,
        '13': 18900,
        '14': 18975,
        '15': 19050,
        '16': 19125,
        '17': 19200,
        '18': 19275,
        '19': 19350,
        '20': 19425,
        '21': 19500,
        '22': 19575,
```

```
        '23': 19650,
        '24': 19725,
        '25': 19800,
        '26': 19875,
}

def send(freq, duration):
        winsound.Beep(freq, duration)

if __name__ == "__main__":
        while True:
                rand = str(random.randint(1, 26))
                send(frequencies[rand], duration)
```