

## Fako Berkers (9651 words)

*I copy pasted my (temporarily) what, how and why into this document and wrote a short abstract. Apart from that I started to write down my thoughts as clear as possible under “Being tracked”. I still have to take out the sharp edges at the end of the argumentation. I think it shows my opinion too much. Right after that part is the rest of my outline followed by loads of annotations from news items that seem relevant.*

### Abstract

This thesis explores how people are being tracked online and why this may be a problem or how this can be justified. My project is an attempt to make clear which risks and changes are involved concerning the huge amount of data that gets generated around our personas in a playful manner.

### What

When people look at the game they will immediately recognize Monopoly. However there is something strange going on. The streets don't represent streets in Atlantis City or any other city in the world. Instead they portray internet platforms like Facebook, Hotmail, Youtube and Twitter. When you buy part of a set you actually buy share from these web services and although you can't buy houses or hotels you're able to buy investments in ad and tracking technology. The “change” and “public funds” cards tell stories about how social media are leaking data about their users and how they are affected by this.

The game exists in two forms. The first is an actual organic game, which has been completely altered to depict a different capitalist world we are used to play with when playing Monopoly, as described above. Another form is a digital form that will be played online. A possible mix of these forms may exist as a performance where I move the pieces by hand as others watch on a distant. These turns are not taken instantly, but as with the abstract version only occur two or three times a week.

### How

For now I limit my description of how to: a table that states how each element in the original Monopoly game will get restyled to fit the new game reality I want to create.

Monopoly	WWWonopoly
Change cards	Messages about data leaks
Public fond cards	Messages about data leaks
Landing on electricity and water company	Draw a change or public fond card
Buying/landing on/trading a street	Buying/landing on/trading shares of a platform
Completing streets of a color	Owning a platform
Buying houses	Investing in ad and track technology (or buying servers!)
Buying hotels	Investing in ad and track technology (or buying a data centre!)
Buying/landing on a railway station	Buying/landing on an internet provider
Get salary	Get salary

Go to prison	Go to prison
Pay taxes	Only two things are sure in life: death and ... ;)
Free parking	Free peer-to-peer downloads
Take a mortgage on property	?

A few rules that are nice to add here:

- There is no such thing as the free p2p download jackpot (as the game would last forever)
- Each player begins with two randomly assigned properties for which they must pay (to speed up play)
- A game will last X number of turns

## Why

It lies in our nature that we can't assert risks very well. A lot of people like social media, but very few seem to be aware what the risks of them are. By allowing players to take the perspective of a “system agent”, meaning somebody with power over the functioning of the social media system, the players get a bird view of what is happening. From this perspective the players are more likely to get aware of risks than when they are “in” the system with a frog like perspective. It would be good if by playing the game and reading the bits of text and articles that are attached to each game message, the player gets a better sense of how social media function and where the vulnerabilities of the system lie. Any system can get played, but it's important that users don't become victims because of that. By informing users in this somewhat playful manner an early warning could prevent harm.

## Being tracked

When I say that you're tracked online I mean that a lot of things that you do when you are using the web is tracked, saved and used by big companies to make money. I will talk about a few ways in which information about your behavior online is valuable and how it finds its way into the hands of who is interested in it. The fact that you are being tracked is problematic when sensitive information is disclosed to parties who were not supposed to have that information. An example of disclosed information is when you want to surprise your partner with a vacation and the surprise is ruined because commercials for the destination you searched for is shown on a lot of online advertisements. Another example is when insurance companies heighten your yearly fee, because you've been searching for AIDS related topics online.

You may think that this kind of disclose doesn't apply to you or is not so bad and that you don't have anything to hide. While this is probably true you can ask yourself whether your friends and family are in the same position and whether you may be in need of privacy as you get older for instance and your medical data becomes more valuable to insurance companies. I hope you come to the same conclusion as me and find that we are better off in a society where we don't have to worry about whether sensitive information gets revealed or not.

I could sum up a set of incentives which would protect your privacy and stop writing this essay. In fact I have made this list here <<<a link here>>> for anybody who doesn't want to read further than this. However a rule without an explanation of why that rule is important is just another rule and such rules are likely to get broken. That's why I want to explain you in simple terms why these rules are important and how they affect your privacy. It will take a little more effort to understand the essay than to read the bullet point list of incentives, but you are more likely to remember and follow rules once you understand their reason. Once you know these rules you can not only protect yourself but also people who are in greater need for it and haven't red this essay themselves.

A website is a collection of webpages. It's best to picture webpages to be some kind of Power Point slides. This metaphor for what a webpage is works well, because like a Power Point slide a webpage can change, for instance when you click somewhere or automatically after some time. A difference between a PowerPoint slide and webpage is that any changes always become visible, while on a webpage this is not necessarily the case. In fact tracking is for a significant part done by changing the webpage invisibly when you are looking at it!

When you visit a webpage on a website you're using a computer program that is called a browser. Examples of browsers are Internet Explorer, Firefox, Chrome, Safari and Opera. As soon as you click on a link somewhere your browser makes connection with a computer which stands in a room like you see below:



Many powerful computers are stored in those things that look like fancy refrigerators on the image. When you see a webpage in your browser it has been send to your browser by one of these computers.

However there is a catch to this that is important for the way that you're being tracked. Your browser receives an entire webpage in parts. The first part, which usually contains all text on the webpage, will also indicate to a browser where it can find other parts. These parts may be images, video or audio that are on the webpage you're visiting. Since these parts are often essential to the look of the webpage your browser will download these parts without checking if they are necessary. The additional media parts could be located on the same computer as the webpage your visiting, but they could also be present on computers which are in a room on the other side of the world as the computer that gives your browser the webpage. A website logo is often retrieved from the same computer as the first part of any webpage which is part of that website. However some images are almost never on the same computer as the webpage, that shows those images. Think for instance about Facebooks "like it" buttons or Youtubes video players, but also banners and commercials. These media are retrieved from Facebook, Google or another media company. When that happens the computers, from these companies, not only provide the requested media, but also record that you have received the media when visiting a particular webpage from a particular website.

Now the question remains how these 3<sup>rd</sup> party computers (computers other than the computer delivering the main part of the webpage) know that it is you who is visiting the webpage. Lets make clear that in principle you reveal as much about yourself to a 3<sup>rd</sup> party computer than to the computer who serves you the main part of a webpage.

Whenever you are asking for a part of a webpage your browser tells a few things to the computer that is supposed to deliver that part to you. It depends on the browser what gets told exactly, but usually it will tell which browser you are using and if you are on a Mac, Windows, Android or other kind of computer. It may also reveal your language setting and which webpage you were visiting before. This information combined already tells a lot about you, because only a few people will send exactly the same information. This makes you identifiable. The really revealing information that gets collected however is the IP address in combination with the content of a so called cookie file.

When your browser connects to a computer for a part of a webpage this computer does need to know where to send the part of the webpage to. To inform websites where you want the webpage parts to be delivered your browser specifies your IP address (Internet Protocol Address). This address that basically functions like your post address is given to you by your internet provider. Any part of a webpage that a browser asks for gets delivered at this address where the browser will put them together to form a whole. Sometimes you share your IP address with the people you live with and it can also happen that you get a partially new address every time you turn on your computer (depends on your internet provider). In general the IP address will limit the size of the group of people who may have asked for the webpage significantly and thus comes very close to identifying you.

The identification is further completed by the use of cookies. Cookies are files that are stored by your browser when you are visiting websites. The content of these files is determined by the computers who send you the parts of a website. So the computers not only send the media or other kind of webpage part you're interested in, but also what the computers want your browser to store in a cookie file. Your browser obeys by default. Whenever you ask again for something from the website, your browser will not only request for the webpage parts, but also remember the website computers what was stored in the cookie file. The content of a cookie file might be the products you've been shopping for so far on a certain website for instance. When you visit the paying webpage the content of the cookie file gets transferred to the website. In this case a correct bill may be generated from this information, since it contains all products you shopped for, which is then shown on the webpage you asked for.

So far so good, but what if a website stores something like a social security number inside a cookie? In that case this identifying number gets send over to the website every time you visit a webpage from that site or download a webpage part from that site. This means that if a certain image or other part of a website is used on lets say a million pages, your travel path between those pages gets tracked precisely. Every time you visit a site from the large collection the same, often useless, web page part is collected from the 3<sup>rd</sup> party computer. Your browser will tell this computer who you are because your identity gets send over like all information stored in a cookie file. The 3<sup>rd</sup> party computer will be programmed to record everything you do within those million webpages and relate the information to the big number given to you.

The technique described above is deployed with Facebooks "like it" button. Since this button is present on so many webpages these days Facebook is enabled to track people over the internet. People don't even need to have an account with Facebook to get tracked, nor do you need to click the buttons in order for it to work. Their sheer presence is enough to tell Facebook you've been visiting a page. The "like it" button part is separately retrieved from a computer owned by Facebook and not the website you were originally visiting. Since every request for a webpage part that needs to be collected from Facebook will be done with your browser sending the information in your Facebook cookie file back to Facebook, the unique number created by Facebook for everyone tells Facebook exactly who you are.

If you register with Facebook that is a bonus. Your name and everything else you share on Facebook will get related to the identifying number too. That way they'll know a lot more about you then where you've been online. The same kind of disclosure happens when you visit webpages with your mobile phone. The browser on your mobile phone can send additional information about you, for instance your Google account ID when you're using an Android phone.

Now let's get back to the idea that a webpage is sort of like a Power Point slide which can change automatically or through interactions, without the visitor noticing. Whenever a website changes a 3<sup>rd</sup> party computer can again be involved for this change and track visitors further. A webpage may at any time connect to a 3<sup>rd</sup> party computer and inform that computer, which actions a visitor has taken. These actions may range from: where a visitor clicks upon, on which part the user is pointing his/her mouse and for how long or how long the visitor is staying in general. This is possible because a webpage is not so much a page, but rather a dynamic collection of different media originating from different sources.

The kind of tracking I've been explaining so far is called web beacons. There are three alternative ways of tracking which are worthwhile to discuss as well. The first type is called analytics and Google Analytics is probably the largest player in this kind of tracking. This service by Google is used by marketers of various companies who want to know which pages of their website get viewed most often, how long visitors are staying and which links they click to leave the page. They use this information to adjust the website and monitor which marketing actions are effective and which aren't. To make Google Analytics work a webdeveloper has to install Google Analytics programming code on each page that needs to be tracked. What Google Analytics does when it is installed is not only save a unique identifier in cookies, but also the time when you requested for the page and which page you requested before. This way graphs about visitors can be generated for marketers by Google, because the information that is necessary to plot such graphs is gathered by the analytics software.

Recently Google has started a new program called Screenwise. It is an opt-in program, meaning that you have to apply for it, and if you do Google will fetch the Google Analytics kind of information from all websites you visit and not only the ones where Google Analytics is installed. In return for opting-in you get 25 dollar. It's possible that in the long run Google will use the algorithms, that they may be devising with the group who opts-in for Screenwise, upon data collected by their Google Analytics program. Their recent change in privacy policy could be a step in this direction. Currently they seem to say that they are not allowed to interpret the data coming from Google Analytics, but by allowing data to be shared among their services they are only one step away from using Google Analytics data in any of their services. Surely Google will try to sell their use of this data as being user friendly as soon as they deem algorithms coming from the Screenwise program a success. However a lot of people are critical about the current widening of allowances for Google so it may be a while before they dare to push it further.

Another way of tracking the behaviour of unknowing computer users is by installing beacons not on the web, but in emails. These beacons are usually images as well, but they can also be special invisible parts of an email. They work the same way as webbeacons in the sense that when the images in the email get viewed they are downloaded by your email program/website from a website. This download is registered by this site and usually your email address gets send along with the request to download the image. The result is that the website is capable of determining who reads their email and who doesn't. This information is used by spammers who will focus their attacks upon people who open the emails. A lot of email programs block this kind of tracking, but they are sometimes still vulnerable to variants of the same principle.

There is another method to track what people are doing online. These days a lot of people don't directly go to a website, but they go to it through Google. For example instead of typing my web address in the browser and directly visit my site, people often go to Google, type my name and then visit my site. If you use the internet this way, which is getting more popular, then Google will know what places you visit. This almost goes without saying! What people don't realize is that Google saves this information, which means this may become public one day. You may not want all things you searched for to become public. Another major problem with going online through search engines like Google is that the search engines will often inform the websites what you've been searching for. It's not so bad when I know you've been searching for my name, but what if a medical information website run

by an insurance company, knows you've been looking for cancer?

Next I will explain how all this data can be used to earn money. This will give an idea about how your data is used and where you could become wary for. As said above a lot of your information is used through something like Google Analytics to analyze the visitors use of a website. In the future this information may get used in different ways, but it's not at the moment. Another way for your information to be used is that spammers may act upon which emails you open and which you don't as explained as well.

The main reason where web beacon information is used for is targeted advertisements. You see this kind of commercials everywhere online these days. If you've ever visited a shoe shop online for instance, chances are great that you get a lot of shoe commercials for weeks to come. Through a web beacon it has been noticed that you visited the shop, but didn't buy anything. To persuade you commercials coming from the same 3<sup>rd</sup> party computers as the web beacons will consist of shoe ads. In that case the web page part consisting of a commercial block will be generated on the fly, especially targeted at you, or at least at the unique number that your browser has sent towards the ad company.

Companies who follow you for this line of business come in three flavors. They all have in common that they get paid by other companies willing to pay for advertisements in the hope that they will have some effect on your buying behavior.

The first kind of company who watches where you are going and creates a profile based on what they now about the content of the webpages you visited. Whenever possible they'll show you ads that in their opinion are related or relevant in relation to the content you've been reading so far. This may be quite far off. If I have a tooth pain and I visited a forum about that I may get commercials about toothbrushes. It has to do with teeth, but it's not the solution I'm searching for probably.

The second kind of company are called retarget companies. These kind of companies are mostly active in and around webshops. They'll monitor your buying behavior and when you don't buy they'll show you commercials from the products you haven't bought yet. Of course it is their hope that this time you'll get persuaded and make a buy online from the company they are working for.

The last variation of company are working like accountants. It would be easy for a retarget company to say that they've been showing all kinds of commercials about products that users didn't buy before, but how does the seller of these products know that this is true? They'd have to track people themselves in order to know that these companies are not conning them in some way. Instead of tracking people themselves they leave it to specialized companies who monitor the monitors by monitoring you.

Advertising is the biggest application of the data that is collected about you at the moment, but there are other uses thinkable which may already be in use albeit in a less open way as advertising.

It will be very interesting for insurance companies to know about which kind of diseases you've been searching for online. This may tell them how big a risk you are for them and what the changes are that you may be knocking on their door for money. If they deem this change very high they may heighten the fee you have to pay them. In this case the data that is collected on the internet becomes another part of your file with these agencies determining whether you'll get cheap insurance or not. This may seem far fetched or unethical to you, but the fact is that insurance companies have done similar things in the past. In the Netherlands for instance insurance companies have payed a freelancer to talk tax officials in giving away social security numbers of people who the companies wanted to check. It wouldn't surprise me if some companies, under false pretenses, were already busy gathering your medical data and selling that to interested parties. There is a lot of money to earn with insurances!

Another application of this data and which is similar to insurance companies is with banks. When you go to a bank to get credit the bank will do everything it can to make sure you're good for your money. Seeing whether you gamble or tend to spend a lot in shops may be an indication for banks

to trust you or not.

The third field where this kind of data can be useful is employment. An employer may want to know more about you and ask a company to generate a report about your online behavior. Before you can say anything about yourself an image about who you are is already made and such first impressions often count.

For all these uses of information counts that the companies asking for the information may not care about whether the data is giving a correct image. It's likely that it is more profitable for them to assume the data is correct and damage some people unjustly than take the risk of losing a lot of money by trusting an entire group they'll know is likely to act against their own interest. You won't get benefit of the doubt and they won't try to talk about it if only because they don't want this sort of tracking practice to be known by the public.

A completely different, but also profitable way of using your online behavior is what some travel agencies have been doing. They measure how often you would look at a certain flight. If you looked often or flew regularly they would heighten the prices for you only, because they make the guess that since you've been thinking about it for a while already or you regularly take that flight you are willing to pay a little more than others.

Of course it is not only companies who can have benefits from this sort of data. An employee from a perfectly behaving company may decide to sell data to interested companies for some extra money. Burglars may pick up on travel plans together with your address and people running political campaigns may be interested in what you think is important to present you a digital pamphlet that addresses all those issues you're worried about.

I'm sure that this long list of possible uses is limited and that there are uses people haven't even thought about. The worrying thing is that this kind of practice is usually well hidden until it goes wrong for the companies and a scandal emerges. Even if this way of doing business is revealed it doesn't always get enough media attention to change things.

### **Reasons why tracking is good**

Of course there are also benefits to being tracked. The most mentioned benefit by Google, but also a lot of other companies and bloggers, is that tracking improves the user experience of advertisements. I disagree with this opinion, but before I explain why I will go into arguments that seem legitimate reasons to track people according to me.

The first reason why tracking is a good thing is that it enables machine learning. Machine learning is somewhat equal to "wisdom of the crowd". How this works is maybe best explained by using an example. Let's take Google's spelling correction. I'm sure everybody misspelled something in Google ones in their life. In the past the phrase "Did you mean:" with the correct spelling of the word you may be looking for was displayed above your search results. Those results were often useless because it wasn't really the thing you were looking for. If you clicked the link you would be brought to results with the correct spelling, without having to correct your own mistakes. Nowadays Google often doesn't even show results with a misspelling. When it is reasonably sure you've made a spelling mistake it will automatically search for the correct spelling and the link above your results will say "Showing results for:" with the correct spelling behind it and a link beneath it saying "Search instead for:" with the incorrect spelling. The computers at Google have "learned" how to spell correctly through the following mechanism. In the beginning there was no spelling correction. In that time people would behave predictably, because when they made a spelling mistake and the results were off, they would correct themselves and immediately search again with the right spelling. As soon as Google started to track and record what people were searching for it must have been obvious that people searching for 'througj' afterwards searched for 'through'. The amount of people that did this could have been in the hundred thousands. From that moment on Google's computers were probably told to suggest another spelling whenever somebody was searching for a word that 1) wasn't in the dictionary

and 2) was often followed by a search for a word that was in the dictionary. The reality is more complicated than the example I give here where somebody mistyped a word, but the principle that I describe where the behavior of many thousands, up to many millions, of people have showed computers what correct and incorrect spelling is holds up. Adding manually all possible spelling mistakes would be super inefficient in comparison to letting a computer learn by allowing it to watch humans. Similar methods are being used with Amazon when they recommend books or LastFM when they recommend music. Books can get recommended, because a computer has “seen” that book A is often bought together with book B and that B rarely gets returned. Important for my argument is that it is unnecessary to know who exactly did what. To go back to Google it is unimportant to know who made the spelling mistakes. The “wisdom” of the computer is created because many people made the mistake, not because a specific person X did it. Identity just doesn't matter for machine learning, but it may matter to employers who want to have a fast way to check whether an applicant for a job is capable of spelling really good and types carefully. Leaked data from Google may indicate this to the employers, who are willing to pay a small price for the comfort of knowing somebody is good or not. As stated before the spelling correction would work without knowing who did what. As a user we could benefit greatly from machine learning without running large privacy risks as mentioned above.

Another reason why tracking could be a good thing is that critical minds can have a look at the data and say something about our society. In other words, the information that gets collected when people are tracked online may be valuable to sociologists or other scientists. An example where the availability of sensitive data to scientist have helped society happened in the previous century. When information about American citizens and house mortgage loans were combined it was shown that black people had to pay more for their house if they wanted to live in a white neighborhood, compared to the white people already living there. When this discrimination was proven by scientists the American government made legislation to make this an illegal practice. I'm not saying that this revelation through sensitive information has got rid of discrimination, but it may have helped. More recently data journalism was done on Facebook. When a Republican was mentioned by name in any kind of message (including private messages) the “mood” of that message would be measured probably by counting the amount of positive and negative words in the message. This “mood” indicator would in turn say something about how the elections for any Republican candidate were going. Such an application of information gathered from people is not nearly as useful as fighting discrimination in my opinion, but it remains a fact that good may come from using the data. The way in which people are informed about the use and a meaningful possibility to withdraw your data from analysis could make practices like this acceptable. In the case of Facebook these conditions were not met unfortunately. Something that Facebook did correctly in my view was making sure that the political website only received the “mood” indicators without knowing who was responsible for which “mood”. It goes without saying that knowing who talked about which Republican in which “mood” would be very valuable to the persons organizing the political campaigns. With sociological research or data journalism the same thing is true as for machine learning: sharing what we've done online can be useful and it doesn't mean we have to be identified in anyway.

A side to tracking that excites me is that in theory the kind of research and tool creation that is described in the previous paragraphs is not limited to big companies or universities. If the data is available through an API everyone can in theory have access to it and use it to develop innovative services. When you use Google, either by typing search terms in the browsers address bar or by typing them on Google directly you are interacting with Google. This interacting happens through a place where you can type your search and a button or enter key on your keyboard. These elements through which you are interacting with Google are called an interface. An Application Programming Interface is a means to interface with an application in a different way than you're used to. You have to see application in a very broad sense. Google, Facebook and Twitter are considered webapplications, but an organized set of data can be called a database application. Whatever the application is an Application



Programming Interface (API) allows programmers to use that application in an automated way. By using the Google API a programmer could automatically search every word of an article in combination with the writer of that article. The program could then tell by looking at the amount of results it gets how often the author has used that word on the web. If you would leave out very common words this may result in an idea about the vocabulary of that author. You can do the same thing manually using normal Google, but interacting through the Google API gives the result in seconds instead of hours. Similarly APIs created for specific sets of data that are gathered by tracking people could be made available. Of course the data should be made anonymous before it is made available like this. At the moment too few people are knowledgeable about the possibilities of interfacing through an API to make the anonymous process and developing of a save API worthwhile. The highest risk for such an API would be that people are re-identified. If an API for an application that tracks grocery shopping habits is made available and name and address are obscured, but peoples general location is not, then there is a good change that the databases of delivery companies are capable to match the weight of delivered packages with groceries done. Since those delivery companies are sure to save your name and address the patterns of doing groceries and the patterns of delivered packages will together reveal who you are. It is a bad idea to let everybody know about your groceries, because insurance companies are already trying to predict your change of getting diabetes based on your daily shopping. If they get this information you may need to pay more to get the same kind of insurance as before.

Another reason why tracking is a good idea is to give marketers and content providers an idea about what is popular and what not. Marketers and content providers always try to predict what is popular and which part of a website will draw enough visitors. Sometimes however they are wrong in their estimates and it is a good thing that the professionals can then adjust their strategies and serve the public better. It would be a bad thing if things that people often come back for are hard to find while things that are almost never clicked are very prominent on the front page of a website. That's bad news for all parties. Similarly people could be tracking you on their blog. This is good for people who are writing that block, because they can go to employers with actual data and say, look I attract so many people on this subject, take me as a content creator or expert on the topic.

<<<burglars??>>>

### **Demystifying the user experience**

A lot of advertising companies say they track people to deliver more interesting ads. If you reason this way, tracking people is a good thing, because it improves the user experience. I agree that some technology as explained above does add to the user experience, but they rarely need to identify you. The big difference with targeted advertisement is that they do need to identify you, which is usually done through the use of cookies, and I don't agree that it makes the user experience any better.

Chris Anderson, a wired journalist, makes a distinction between hit and tail products. In the past we would only be focused upon hit products. These are products that are popular with a lot of people. The term hit comes from music hits. Just like Madonna draws a lot of public and sells large quantities of music to them, so are there some products like Coca Cola who are well known and sell very good. Tail products are different from hit products in the sense that they are not very popular and do not sell very well, but because there are so many more tail products than hit products they all together sell better than the hit products. The example that Anderson gives is with music and a digital service called Rhapsody where you can buy individual songs. Rhapsody sells a lot of Madonna and a lot from popular hit artists like her. However the revenue that they get from these artists is only one fourth of their total income. The rest come from long tail artist who do not sell a lot, maybe as less as three songs per year, but because the tail artist are with hundreds of thousands they together sell much more than the handful of hit artists combined. This long tail kind of market is very suitable for digital markets, because practically all content out there can be made available online, which makes the tail longer. There is no issue with making everything available because the costs for making any artist available is very low

and can almost be considered to be zero. Bloggers and mouth-to-mouth commercials point consumers into the right direction along the tail. As explained by Anderson the existence of computers who have learned what people like are also important. There is no way for a human to go through all the content. Instead a computer can make recommendations and the best way to do this is through machine learning which involves tracking people as I explained earlier. However this kind of tracking can be done completely anonymous since there is no need to know who bought what, but only how often product A was purchased in combination with product B. The identity of the buyer is irrelevant. So let me recuperate. Companies make huge amount of money by offering not what is best known, but everything available. They can do this only because they work on a huge scale and a lot of small sells add up to be a lot of money. The guidance of consumers are first of all bloggers and friends who point to the right direction, but secondly purchases on these “long tail platforms” are tracked in order for a computer to do recommendations based on what other customers have bought before.

It's clear that companies like Amazon and Rhapsody are delivering an interesting user experience and I have used their services myself. In their nature they are however completely different from targeted advertising. Targeted advertising always has to do with a limited amount of suppliers. If all suppliers were automatically listed with one advertisement, why would advertisers pay for the service Google provides? Google makes money, because it only allows those who pay. The more popular a search term is the more money you have to pay for it. This means that on a lot of terms you will mostly see advertisement which are in fact hits asking for more attention, because only those products individually acquire enough money to justify an investment in advertisement. The further you go down the graph towards the tail the more interesting and surprising products become. We need very strong filters to get rid of the noise of products that we don't need, but targeting advertisement is not the right kind of filter, because the further you go down the tail the less money will be available for advertising. Remember that those products wouldn't be available at all if it wasn't so cheap to make them available. Adding cost to them by investing in advertising will certainly make it less profitable if not downright a loss.

It's true that if you go down the line and use less popular search terms you don't spend a lot of money and you won't lose a lot of money with buying a commercial. The problem with this is that it is then necessary to use very specific search terms. Using such search terms partially defeats the purpose of advertising, because your product will already be known by the consumer if he/she is able to write down those search terms in a search box. <<<bloggers article>>>

There are clear indicators that targeted advertising is in a completely different realm than the user friendly experience we have with other digital platforms. This suspicion is confirmed when you look at the click-through-rate of commercials on Google. This percentage stands for the amount people have clicked on an ad and went to read more about it. The average click-through-rate is a shaming 2%. If only 2% of the users is interested in it I wouldn't dare to call it a user experience success. You may wonder how Google is able to make a profit with such a low success rate. The answer is again scale. Google is showing those commercials to so many people that 2% ends up being a lot of people clicking it.

I can imagine that commercials on highly specialized technical blogs are considered to be an improvement for the visitors. The technological specs of a device are very specialized and the cost of such devices are usually huge anyway so spending a bit on digital advertisement won't be a problem. I can imagine that users mistake such ads for content as Anderson describes in his book, but technological products for hobby use are a very biased category. What about shoes? It will be impossible to list a search term about shoes that visitors will use and at the same time will not cost a lot of money. Most advertising of products will simply fall into the hit category and cost more than products in the tail could ever afford. That is if they are even available online in a tail sort of fashion, because a lot of products still aren't! <<<what if the long tail algorithm is extended into an ad?? only the huge companies are able to do that>>>

Another category of online advertisements are the re-target adds. These adds will appear when a computer thinks you were about to buy shoes, but eventually didn't. I don't think this kind of advertisement can even be considered to be user friendly. For whatever reason the visitor decided not to buy, the change is not very big that the visitor completely forgot about the product and is grateful for the reminder. It's more likely that people who were first able to resist return to the product when they see it again, but I would rather call that manipulative than helpful.

The final category of advertisements that I want to discuss are social advertisements. These kind of advertisements appear on networks like Facebook. It will display products that your friends were interested in or even bought and the advertisements will usually display who it is. If they are real friends and truly think their purchase is worth mentioning they will do it themselves and don't need computers to mediate the information. When they are more like acquaintances which are most people on my Facebook friends list, then I don't think the recommendation will be important for me in the choice of products. I realize that this is my opinion and it may be different for others. However Facebook has to deal with the same problem as Google, which is that mostly hit products will be able to pay for relevant terms. It has leaked that Facebook is struggling with getting enough income for commercials. This is an indicator that their advertisements are also not so much of an user experience as the platform itself. It has to be said that Facebook doesn't really say it's giving a positive experience with its advertisements. In fact it is rumored that they are busy with doing more intrusive advertisements.

All in all I can only conclude that tracking people online to give them a better experience is a bogus argument. The user experience given by the advertisements fade in comparison to the user experience the platform gives. This is another argument to say that tracking is good, because it allows us to use services for free. I'll get into that next.

### **Paying for services through losing your privacy**

Contrary to popular believe a lot of things on the internet are not free. Distribution online has become so cheap that a lot of companies will not charge you for the cost. These costs are so low, because every year computational power, digital storage and the costs for sending digital information over a network have halved or more than halved every year for the last decades. A problem here is that to buy these resources cheaply you'll have to buy a lot. You can get a transistor (a measure for computational power) for 0.000000001 cents a piece, but only if you buy 5000000000 at a time. <<<insert right numbers>>> If you want to buy one they are a dollar a piece. This means that if you want to provide your customers with free computing power like a free filter algorithm, you'll have to invest greatly. The alternative is that you let your visitors wait. This is exactly what I did with my own filter algorithm for Wikipedia called Wicked Wiki. It can take a few minutes before that algorithm is done. If a normal machine takes minutes to search and filter a relatively small part of Wikipedia you can imagine how strong the computers of Google are when they search the entire web in less than a second. It's true that I could do major improvements on my own algorithm making it faster without additional hardware costs, but I'm giving this example to give somewhat an idea of the amount of equipment Google needs. It goes without saying that this equipment is not free in fact it is not even cheap. Google spends hundreds of millions of dollars per year on equipment. The reason why we can use it for free is that so many people use it that the cost per person is so close to zero it isn't profitable to collect it. However the hundreds of millions have to come from somewhere. This is where advertisement comes in. The equipment is payed for with advertisement income. There is nothing new about advertisement paying completely or partially for services. Magazines cost a lot more to produce than the income a magazine gets from subscribers. Advertisement covers the rest of the costs. There is nothing wrong with this model and the money earned with it is estimated to be around 300 billion a year. For this model tracking is unnecessary. Google CEO Eric Schmidt is estimating that the market for targeted advertising is 800 billion. For this kind of advertising tracking is a necessity, because if you're not

tracked, you can't possibly be targeted. In this kind of model we pay not only by seeing ads, but by giving up our privacy. There are a few reasons why our privacy could cost us a lot of money. <<<link here>>>, yet we are giving it away for free. We could charge money, because cookies a vital part of the tracking technology to identify who we are is under our control. Despite that Google is paying some people a few dollars to loose their privacy, this has not yet become practice for the masses. This is a strange thing, because even though the distribution of cookies to Google is virtually costless the information is rare and monetizable. That should raise the price for this information considerably! At least this is what is happening with big media companies who are trying to make money for their rare content in any way they can. I think that it is waiting for a solution for the masses to monetize their rare content somehow.

There is another model that I would like to discuss which could support free web services that nobody wants to loose without the need for tracking. This is a variant upon the freemium model, which is described by Chris Anderson. In this model 5% of the users support the rest of users. In this model there are two products, one is free for all and for the other users have to pay a price, but it works better in some way than the free versions. It turns out that about 5% of the users will be using the advanced non free version. If this version is priced in such a way that a profit can be made for the entire product then effectively 5% of the users is supporting a free version for the other 95%. Often this model works because some people have more time than money while others have more money than time. In a game for instance, where you have to travel in the game world, people are offered to buy teleportation stones. These "magic" stones, which you can buy in an online shop, will get you anywhere in the game in an instance. If you would walk there it could take up to 10 minutes. Some people want to save time and buy stones, while some want to save money and spend time. A lot more people will spend time, but when the price of stones and products like it is correct the game can make a profit through it. Google could make everybody wait ten seconds unless they are a paying user. The people paying would also pay the equipment for users that do have to wait for their results. Similarly Google could have people pay for guaranteed search privacy. The yearly fee for this does not have to be huge, because the ad income per person is very low.

### **Ways to intervene in the technology**

Now that we know what kind of information about you is known and who is trying to make use of it it is time to go into ways of preventing this kind of privacy breaches. There are a few things which I won't get into, because I can't address everything. I do want to mention them here quickly, because I don't want to give the idea that you're save when you do all the things I describe here. The tips that follow now can be read upon extensively from other sources. Where possible I'll try to refer to sources.

The main thing to do is keeping the software on your computer up-to-date. You need to do this because viruses won't get a change to infect your system. The problem of viruses is not really privacy specific, but viruses may try to steal personal information like your creditcard number of bank login password. Every piece of software should be kept up-to-date, but really vital parts are your operation system (like Windows) and two other programs called Flash and Java. Your computer will try to update all these things by itself, but there are a few ways in which you can obstruct this process, without realizing it. The first way is by never shutting down your computer. This has become a popular way of dealing with your computer, because you never have to wait for starting computers. However it is during the startup and shutdowns that updates become effective and viruses are stopt. Up to date virus scanners won't harm you computer, but don't use two, because they may not work well together and slow down your system.

It's possible to observe the internet traffic. This means that people who have the know how can see which sites have been visited by you. The difference between this kind of observation and tracking through web beacons is that observation can not have influence upon what you're about to see. With other words somebody can observe you've been visiting Twitter, but there is no way for these people to

act upon that and change your Twitter experience in any way. This is unlike tracking where often advertisements or search results are changed. However when you send your user name and password or creditcard information over the internet then this can be seen through observation and that information could of course be useful for criminal acts. A lot of people already know that with internet banking you've to make sure that the part on the left side of the address bar should be green. This indicates that the internet traffic has been encrypted. Then it's still possible for people to observe, but it's hard to extract information because all they see is seemingly senseless text which needs to be decrypted before anybody can read it. Recently traffic from a website that Dutch citizens use to identify themselves with the government could be decrypted and that site had to be taken out of service because of it. In general it's important to make sure that private information stays private when it is traveling from computer to computer and your browser will indicate if this is the case.

A variant to the same problem from the previous paragraph is that emails can be read when traveling from computer to computer just as login information and webpages can be observed. This is why it's never a good idea to send confidential information through email unless you use a secure connection with encryption. However with email it is a lot harder to force your email program to use encryption. Partially because a lot of receivers won't be able to decrypt the message and that defeats the purpose of sending the email in the first place!

The ways that I've described above will protect you from criminals that try to get to your personal information. However big companies that know exactly what you do may be a bigger thread to you, than computer criminals. To protect yourself from them you need to do more than taking into account the general computer safety rules.

It starts with picking a good browser. As I said before the browser is the computer program you use to go online. Examples of browsers are Internet Explorer, Chrome, Safari, Firefox and Opera. Most people use Internet Explorer, but this is not the safest browser. The biggest downside is that Microsoft may want to make money some day by allowing companies to track you. They may also track you through their own services like Hotmail if you use Internet Explorer. Safari is the browser which gets delivered with Mac computers and I have the same objections against it as Internet Explorer. Macintosh and Microsoft can't really be trusted to completely stop tracking. Chrome is probably one of the worst browsers for privacy, because it directly goes against Google's business plan to grant you your privacy. They make money by knowing as good as possible who you are! Any browser coming from this company is very hard to trust in my opinion. The best browsers available for privacy are Firefox and Opera and I prefer Firefox because it works slightly better at blocking web beacons than Opera. I'll be discussing how to setup your browser safely using Firefox as an example because that will yield the best results for you.

<<<Split up to different platforms>>>

Unfortunately no browser that I know of protects your privacy optimal from the very beginning. The protection Firefox offers is even less than for instance Safari. You specifically have to tell Firefox that you want to be protected. The good news here is that it is not a lot of work. The bad news is that it needs to be done for every computer separately.

The first step is to disallow third party cookies. Remember that cookies are used by websites to store information on your computer. Often they store login information, but they can also be used to store large numbers that get used to identify you. Third party cookies are cookies that can identify you on a large number of websites and which will link all your visits to these sites together. In the Firefox menu go to Edit → Preferences. A window will appear and you need to click on the tab privacy. Under history you can change how Firefox will save your browsing history including cookie files. Next to the label "Firefox will:" select "Use custom settings for history" in the drop down menu. Make sure that the option "Always use private browsing mode" is NOT selected. This mode is also called the porno mode and is really only useful to keep people who live with you to know what you're browsing for. At

any time you may start this mode by pressing CTRL+SHIFT+P simultaneously. Apart from when looking for porn you may want to use this mode when planning surprise vacations. Lets return to disabling third party cookies. Make user that the option "Accept cookies from sites" is selected. You need to be an expert user to disable that function unfortunately. This is because cookies also regulate the sign in process to websites. If you don't select this option you may find that you're unable to use a lot of services, because logging in is impossible. The option we really don't want selected is "Accept third-party cookies". If you don't select this option it will be much harder for companies to track your whereabouts online! There is however a small price to this. In rare cases login is done through a third party service. Think about login in on a site with your Facebook account, but also Hotmail uses a third party cookie file to keep you logged in (for unclear reasons). I would advice not to login to sites with your Facebook account. To fix login in to Hotmail you can do the following. <<<explain how to add exception separately>>>. Remember that other login services may break and they may not tell you that the problem is that you don't allow third party cookie files. A real setback for privacy, but try to remember that this may be the problem whenever you can't login repeatedly. However the cases where a third party cookie is used for login is very rare!

The last option you need to set regarding cookies is that they should "expire" as soon as you close the browser. After the label "keep until" select the option "I close Firefox" in the drop down menu. This means that any cookie that does get installed will be removed when you close Firefox. If you don't do this cookies are practically stored forever. When you have set preferences for Youtube, Google or some other web services these preferences will be lost whenever you delete all cookies. If this is a disaster for you, you shouldn't do it! Remember that it is not a good idea to keep programs running forever. When you don't need to be online anymore you should close your browser. Not only will this delete cookies, which makes it harder to identify you, it will also make sure that any updates take effect. As explained above updates are important to keep your computer virus free. Updates may be annoying, but they are vital for your PC health. This is why it is a bad idea to suspend your computer to often. If you don't turn off your computer updates may be on hold.

Without third party cookie files the web beacons will have a hard time identifying who is visiting the website. This is already a big step in the right direction! It would be ideal if we could stop the web beacons from working all together and fortunately we can. By installing additional browser software called Ghostery all web beacons will be stopped and you can get notified about their presence. It's quite scary when 12 or more beacons suddenly get disabled on sites where you've been coming for years! <<<explain how to install Ghostery>>>

Next you need to add a search engine to Firefox which will not give away or store your search history. I advice to use DuckDuckGo. You can go to this page and click on the button that says <<<install search engine in Firefox>>>. Then on the right side at the top you can select DuckDuckGo from the drop down list next to the search input field.

Since 2010 it's possible to enable a "Do Not Track" option in Firefox and the <<<NSI>>> has created a system to opt-out of targeted advertising. Both privacy tools give a false feeling of protection and I wouldn't advise to use them, because their effect is so limited. When you enable the "Do Not Track" option your browser will tell all internet computers owned by anyone that you don't want that computer to follow you. This sounds like a great feature, but companies are not obliged to stop tracking. There is a great risk that you feel save, but companies still track you. In fact most companies will interpret your request as for not being tracked as a request for opting out on targeted advertising. This is not the same thing as not being tracked. When you opt-out for targeted advertising you indicate that you don't want to receive targeted advertising. Only normal advertising will be shown from that point on, but you may still be tracked. The data will not be used to give personalized advertising, but its still being collected and you run the risk that one day it will leak into the public domain,

Even if you're willing to do everything that is on the list above. It's still not a good idea to post everything on Facebook. The more information you disclose about yourself to friends the more

information you disclose to everyone interested in you for financial reasons. Everything that I explain here is a precaution to limit the amount of information you reveal about yourself. When you give all this information away because you find you have to share on Facebook instead of telling people personally this would be like closing all doors and windows with locks to keep people out while you forget that your house only has three walls. Please keep in mind that Facebook and Google through email have bad reputations when it comes to keeping your content only available to you. There are a few cases where information seamlessly became available to everyone. Unfortunately there is no way to prevent this except using alternatives. It's understandable that you're reluctant to do this, but try to keep in mind that people may be reading your stuff without being your friend.

A final thing you can do is call your internet provider and ask them whether they deliver “dynamic IP addresses” or “static IP addresses”. When your IP address is “static” you could ask if they also deliver “dynamic” addresses, because you would want these ideally.

Argument against cumulation of capital

- The Mayfair Set by Adam Curtis
- Dark Side of Google (lack of innovation is happening)

Routing leaks?

By about a hundred professionals who occupy themselves with how to score high in Google it is estimated that the use of the website as monitored by Google will become more important. This means that pages which are used intensively will rank higher than they do today. For this mechanic to work it is necessary that the behavioral data is recorded.

<http://www.seomoz.org/article/search-ranking-factors>

Since 2003 Google has started to work with targeted advertisements. In their news item they claim that advertisers and partners get a high return on investment. The weird thing is that they also claim that users get a high return on investment. This bares the question what a user is actually investing and how users are rewarded for this investment.

Google seems to make the argument that if Google makes a lot of money, because there is a high click-through-rate the users must be rewarded for their investment of their privacy as well. This is an invalid argument, because the big profit can also derive from the fact that Google works on a huge scale. If only 10% of the people click through Google's profit must already be enormous. But it actually means that the percentage of users happy with targeted advertisements is quite low while they run the risk of a privacy breach.

<http://www.google.com/press/pressrel/advertising.html>

When Google started to combine information from different profiles there was reason for some to look for alternatives. There are a few things to think about when protecting your privacy:

- search engine
- email
- browser (search engines in browser)
- web beacon scripts

[http://www.security.nl/artikel/40063/1/Een\\_leven\\_zonder\\_Google.html](http://www.security.nl/artikel/40063/1/Een_leven_zonder_Google.html)

[http://www.security.nl/nieuwsbrief/artikel/188/40035?](http://www.security.nl/nieuwsbrief/artikel/188/40035?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/188/40035?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

Googles new privacy policy doesn't change much for Google Analytics. They still won't share visitors data with other parties or even other Google services. However they are busy with a program that seems to be an experiment to use Google Analytics data in the same way they use other gathered information.

<http://analytics.blogspot.com/2012/01/googles-updated-privacy-policy-what-it.html>

It's not very important to focus upon something like fingerprinting. I say this, because 97% of the data leaks that occur on the internet are caused by SQL-injection. SQL-injection is an outdated principle, but it seems that it is still in use vigorously. There are even people pleading to stop trying to secure other kind of leaks before SQL leaks are solved. The same can be said in regard to cookies and browser fingerprinting. Cookies are an old technology, but as long as they are mostly responsible for giving identity away I say we should focus upon them.

<http://www.cio.co.uk/news/3331490/barclaycard-97-per-cent-of-data-breaches-due-sql-injection/>

Google has launched a new service where you can search for symptoms and then Google will do suggestions about the kind of disease you have. Ads are not personalized based on this information and it's possible to wipe your search history, but it remains unclear if websites can pick up upon search terms.

[http://www.readwriteweb.com/archives/why\\_google\\_didnt\\_build\\_search\\_plus\\_your\\_body.php](http://www.readwriteweb.com/archives/why_google_didnt_build_search_plus_your_body.php)



If Republican presidential candidates are mentioned by name on Facebook a company will automatically measure the mood of the message. This is seen as interesting sociological information by some and a privacy nightmare by others. The biggest privacy objection is that there is no way to turn this feature off. An argument for this kind of action is that sociological data in the past has helped to fight discrimination.

[http://www.readwriteweb.com/archives/why\\_facebooks\\_data\\_sharing\\_matters.php?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+readwriteweb+%28ReadWriteWeb%29](http://www.readwriteweb.com/archives/why_facebooks_data_sharing_matters.php?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+readwriteweb+%28ReadWriteWeb%29)

With frictionless sharing, Facebook has changed sharing into archiving peoples consumer habits. Everything you see and read (or actually ... click on) will be put into your Facebook news feed. The author of the source has a strange double attitude towards it. I think it's wrong to see me as a consuming machine with billboard properties.

[http://www.readwriteweb.com/archives/facebook\\_hasnt\\_ruined\\_sharing\\_its\\_just\\_re-defined\\_it.php](http://www.readwriteweb.com/archives/facebook_hasnt_ruined_sharing_its_just_re-defined_it.php)

Google started a big “privacy” campaign about internet. However it doesn't seem to be addressing the issues I'm worrying about in relation to Google.

[http://www.security.nl/nieuwsbrief/artikel/187/39931?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/187/39931?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

It was shown by research that a lot of people do not understand privacy tools such as opt-out, Ad Block, Ghostery and others. They were less well protected than thought or weren't protected at all. Filters were the hardest to understand. The things people run into are familiar. It is a sign that there is still a lot of opportunity to improve privacy tools.

[http://www.security.nl/nieuwsbrief/artikel/175/39042?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/175/39042?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)  
<http://www.sciencedaily.com/releases/2011/10/111031120249.htm>

You can now opt-out for Google targeted advertisement.

[http://www.security.nl/nieuwsbrief/artikel/175/39044?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/175/39044?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

Google admits that its business model is in conflict with privacy, Encryption is no option since Google makes money by selling targeted advertisements and this is impossible if Google can't get to the data.

[http://www.security.nl/nieuwsbrief/artikel/175/39067?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/175/39067?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)  
<http://www.businessinsider.com/googles-business-model-is-in-conflict-with-your-privacy-2011-11>

Facebook is having trouble to make profit. Advertisements are being more ignored than anticipated. This is why Facebook will take a number of actions to make ads more invasive to the user experience.

<http://www.privco.com/recently-leaked-facebook-documents-show-facebook-scrambling-for-improved-ad-performance-to-boost-short-term-revenues-for-ipo-company-trailing-revenue-plan>  
<http://venturebeat.com/2012/02/23/facebook-q1-ad-revenue/>

Users are expecting miracles from the Do Not Track functionality. In truth the companies that will take into account the header will not show targeted commercials, but any tracking could still happen in the background.

[http://www.security.nl/nieuwsbrief/artikel/192/40505?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/192/40505?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[http://www.pcworld.com/article/250689/most\\_consumers\\_clueless\\_about\\_do\\_not\\_track\\_technology.html](http://www.pcworld.com/article/250689/most_consumers_clueless_about_do_not_track_technology.html)

The new privacy policy of Google especially hits Android users, because to use that device effectively you'll need an account and everything you do can now be linked to this device. Canadian officials are concerned.

[http://www.security.nl/nieuwsbrief/artikel/192/40503?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/192/40503?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

Google in trouble because it circumvented default privacy settings of both Safari and Internet Explorer. Google followed them by mistake according to Google.

[http://www.security.nl/nieuwsbrief/artikel/192/40426?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/192/40426?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)  
[http://www.security.nl/nieuwsbrief/artikel/192/40416?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/192/40416?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)  
[http://online.wsj.com/article\\_email/SB10001424052970204880404577225380456599176-1MyQjAxMTAyMDEwNjExNDYyWj.html](http://online.wsj.com/article_email/SB10001424052970204880404577225380456599176-1MyQjAxMTAyMDEwNjExNDYyWj.html)

Google is paying people to join in two programs that are more intrusive than their current ways of tracking people online. The amount of data gathered is similar to what Google Analytics gathers and in one of the cases it is even more than that!

<http://www.neowin.net/news/google-paying-people-to-track-their-web-visits>  
[http://www.security.nl/artikel/40256/1/Google\\_betaalt\\_internetter\\_20\\_euro\\_voor\\_privacy.html](http://www.security.nl/artikel/40256/1/Google_betaalt_internetter_20_euro_voor_privacy.html)

Chrome users are being followed to detect malicious phishing sites and downloads. IP addresses are gathered, but are stripped from the URL after two weeks.

[http://www.security.nl/artikel/40141/Google\\_gebruikt\\_Chrome-gebruikers\\_als\\_honeypot.html](http://www.security.nl/artikel/40141/Google_gebruikt_Chrome-gebruikers_als_honeypot.html)

According to some scientists it is very hard to de-identify people. Presumably this is because two data sets can be linked together and make the group of possible matches a lot smaller. However according to this article it is hard, but not impossible to de-identify. It is good to do this kind of thing, because it may help public health if data is looked at.

<http://www.sciencedaily.com/releases/2011/06/110616092650.htm>

Twitter was uploading user contacts without informing them. They admitted this and you can now remove the data, the question remains how many people know about this? Another startup called Path did a similar thing.

[http://www.readwriteweb.com/archives/twitter\\_is\\_the\\_latest\\_company\\_to\\_admit\\_it\\_uploads.php](http://www.readwriteweb.com/archives/twitter_is_the_latest_company_to_admit_it_uploads.php)

A burglar was using Google Maps to scan the area before he struck. It's possible to remove your house from Google Maps.

<http://www.businessinsider.com/how-to-protect-your-home-from-a-google-maps-burglar-2011-9>

The new privacy policy of Google is there to allow Gmail related data to be used in Youtube and vice versa. This is probably done to get more income from advertisement.

<http://www.businessinsider.com/heres-the-crucial-part-of-googles-new-privacy-policy-that-has-advertisers-salivating-2012-1>

Googles side of the story (concerning 3<sup>rd</sup> party cookies in Safari) is that people opted-in for +1 buttons

on advertisements. Somehow Doubleclick also got access through the loophole made for these +1 buttons.

What we can also conclude from this is that Safari doesn't block anything as long as users once trusted the sites from trackers. This makes from a moderate privacy protection I would say.

<http://www.businessinsider.com/google-apple-tracking-explanation-2012-2>

When you use an Android phone and ads appear in your browser. Your phone number is connected to the behavioral data according to this source. However the source doesn't think this is a risk in anyway and again the argument for beneficial advertisements show up.

<http://www.businessinsider.com/why-you-want-google-to-know-everything-about-you-2012-2>

A science fiction story at the moment, but it may be that salespersons or people giving away flyers will be prompted to reach you based on your Google settings.

<http://www.businessinsider.com/how-google-apps-may-soon-let-salespeople-stalk-you-2012-2>

A very violent text about how privacy doesn't matter anymore.

<http://www.businessinsider.com/online-privacy-who-cares-2012-2>

A very good text about how privacy does matter, but is a complicated whole. It tells a little about the companies that are following you!!!

<http://www.theatlantic.com/technology/archive/2012/02/im-being-followed-how-google-and-104-other-companies-are-tracking-me-on-the-web/253758/>

More official source about how web beacons work.

<http://priv3.icsi.berkeley.edu/>

By leaking a large amount of personal data it's possible that users become victim of identity theft.

Loosing this data also helps to de-anonimize other databases. It has been shown that such data can do that.

<http://www.sciencedaily.com/releases/2012/01/120118122829.htm>

It may be a good idea to split up different data sets about oneself for different purposes.

<http://www.sciencedaily.com/releases/2008/12/081204094555.htm>

75% of websites leak data to third parties. 50% of the websites leaks identifying numbers, which can be cross matched to identify somebody. It's impossible to say which companies are bad and which aren't. In the mean time it is clear that there are high incentives to collect data and that those companies can't be trusted to protect privacy. The responsibility should lie with 1<sup>st</sup> parties according to the article. All the consumer tools available leave some kind of hole in the security according to this report.

<http://www.sciencedaily.com/releases/2011/06/110602111437.htm>

Data about Google employees was stolen (through burglary) which can be used to perform identity theft and make credit card accounts on behalf of others.

[http://news.cnet.com/Stolen-Google-employees-personal-data/2100-1029\\_3-6243093.html](http://news.cnet.com/Stolen-Google-employees-personal-data/2100-1029_3-6243093.html)

Both Facebook and Google have fired people, because employees were accessing personal data from users that they shouldn't look into. According to Google it has only happened twice in 10 years time with 20.000 people working there. A very low number which may be false, but Facebook isn't commenting at all. Apparently somebody who was working at the help desk also misused his

knowledge power and harassed teens.

<http://techcrunch.com/2010/09/14/google-engineer-fired-security/>

<http://thenextweb.com/google/2010/09/14/ex-google-employee-dug-through-private-data-and-harassed-teens/>

If you searched for herpes, then the search term together with your IP address and possibly a cookie will be recorded. This could lead to unwanted things with acquiring credit and health insurances.

[http://www.security.nl/nieuwsbrief/artikel/187/39911?](http://www.security.nl/nieuwsbrief/artikel/187/39911?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/187/39911?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

There is a plugin for Firefox that enables social media plugins, but just doesn't send along any identifying cookies, when such plugins are loaded.

[http://www.security.nl/nieuwsbrief/artikel/171/38759?](http://www.security.nl/nieuwsbrief/artikel/171/38759?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/171/38759?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

Possibly 10.000 Facebook accounts stolen, with the remark that when people use weak passwords they may use the same password for their email and the email account is a gateway to other passwords which are often changed through email.

<http://countermeasures.trendmicro.eu/over-10000-facebook-account-details-hacked-and-published/>

Google engineers are saying that cookies are like supermarket discount cards. However I never received any thing more than free. And the discount cards are not necessarily a good practice.

[http://www.security.nl/nieuwsbrief/artikel/173/38867?](http://www.security.nl/nieuwsbrief/artikel/173/38867?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/173/38867?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

The EU is busy with legislation that forces companies like Facebook to keep data from EU citizens save regardless of the location of the servers. The legislation seems to include the right to be forgotten,

[http://www.security.nl/nieuwsbrief/artikel/173/38867?](http://www.security.nl/nieuwsbrief/artikel/173/38867?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/173/38867?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[http://www.security.nl/nieuwsbrief/artikel/188/40024?](http://www.security.nl/nieuwsbrief/artikel/188/40024?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/188/40024?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[http://www.security.nl/nieuwsbrief/artikel/188/40045?](http://www.security.nl/nieuwsbrief/artikel/188/40045?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/188/40045?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

A man was taken out of his car insurance, because he was present at a race car event. The man is proceduring. It's clear that insurance companies are using Hyves and Facebook to look for what they see as fraude.

[http://www.security.nl/nieuwsbrief/artikel/176/39090?](http://www.security.nl/nieuwsbrief/artikel/176/39090?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/176/39090?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

The Facebook user is unsure if Facebook sells data to third party companies and how to protect themselves against this.

[http://www.security.nl/nieuwsbrief/artikel/174/38948?](http://www.security.nl/nieuwsbrief/artikel/174/38948?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/174/38948?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

More than halve of the most visited websites on internet leak information to third parties. This includes usernames.

[http://www.security.nl/nieuwsbrief/artikel/172/38817?](http://www.security.nl/nieuwsbrief/artikel/172/38817?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

[utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=newsletter](http://www.security.nl/nieuwsbrief/artikel/172/38817?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter)

500.000 British citizens were being sold by corrupt Indian call center workers. The data included creditcards and sensitive information about mortgages and loans. Companies are unwilling to work with authorities, because they fear a scandal.

<http://www.dailymail.co.uk/news/article-2116649/Indian-centres-selling-YOUR-credit-card-details-medical-records-just-2p.html>

American teens were fined, because they befriended a beautiful girl online, who was actually a policeman. The teens were caught drinking underage.

[http://lacrossetribune.com/news/local/article\\_0ff40f7a-d4d1-11de-afb3-001cc4c002e0.html](http://lacrossetribune.com/news/local/article_0ff40f7a-d4d1-11de-afb3-001cc4c002e0.html)

There is a list online that says what Google is gathering in personal data.

<http://www.seomoz.org/blog/the-evil-side-of-google-exploring-googles-user-data-collection#list>

By blogging you'll use long tail search terms. This will attract attention towards you.

<http://www.hodgesnet.co.uk/?s=long+tail>

Microsoft and Yahoo struggling to get ad revenue

<http://www.marketingpilgrim.com/2009/04/microsofts-online-advertising-revenue-drops-16-time-to-give-up.html> (old)

[http://www.pcworld.com/businesscenter/article/236036/yahoo\\_revenue\\_drops\\_as\\_display\\_ad\\_business\\_slows.html](http://www.pcworld.com/businesscenter/article/236036/yahoo_revenue_drops_as_display_ad_business_slows.html)

11 weeks

First sketch of entire argument in week 16 (4)

Week 17 for teachers to respond (1)

Mapping of argument to the project in week 18 (1)

Reading of entire paper by in week 19 (1)

Rewriting in week 20 (1)

Rereading in week 21 (1)

Final writing in week 22 (1)