HOW MUCH DATA ARE PEOPLE GIVING AWAY?

This paper aims to outline the technologies being used to gather data about individuals, how this is being used and what the future could hold for data gathering.

INTRODUCTION

In today's connected world more and more users are storing information about themselves on the Internet and within social media websites but also the amount of information available about an individual has also increased exponentially. Many users are oblivious that they are even sharing their data; others simply do not understand the security risks this poses. With the penetration of smartphones and other connected personal devices into the market it is also possible to track an individual's movements. In this paper we will highlight the various ways data about a user is available and how this can be and is used today.

AN INDIVIDUAL'S DATA

Every minute of every day, millions of users update social networking websites with what is going on with their lives, sharing it with their friends, but **how many also realise that this is shared with the owners of the website?** In December 2010 while researching for his thesis, Max Schrems requested a copy of all the information that Facebook held about him. Facebook replied with a 1200 page PDF document which highlighted such things as his last location taken from a combination of "check-in's", IP addresses or geo-tagged uploads (Solon, 2012). Schrems searched the 1200 page document to highlight all the categories in which Facebook stores information. These can be found in Figure 1 - Wired.co.uk below:

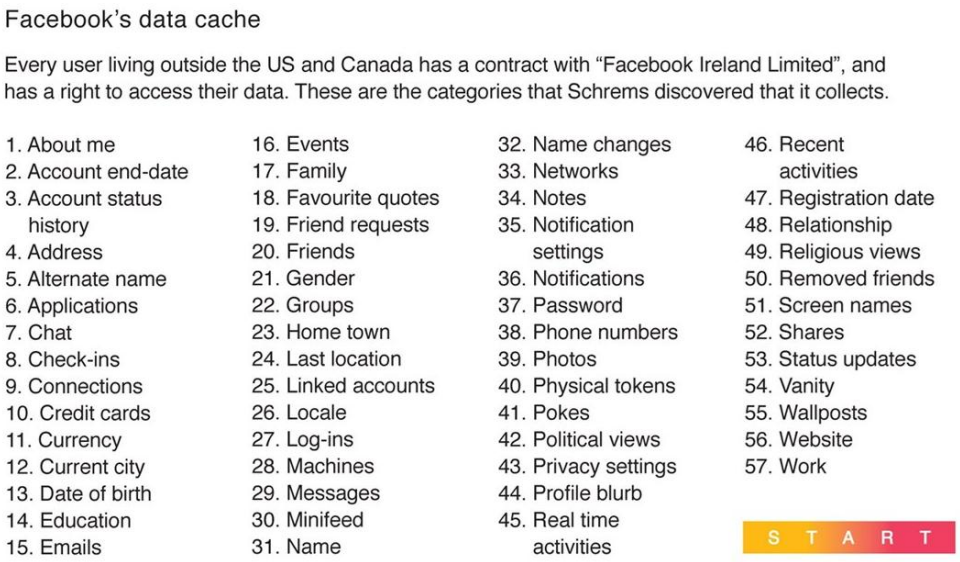

Figure 1 - Wired.co.uk (Solon, 2012)

Facebook is not the only technology giant actively tracking and monitoring behaviours. Google is not only the world's largest search engine, its technologies provide users a variety of different services:

- email (being one of the top 3 email providers in world),
- a social network,
- a blogging platform,
- the world's largest video sharing website,
- the Android phone and
- Google Chrome the most popular web browser in use today.

With all these at its disposal, Google is well placed to find a wide range of information about any single person. From Google's Data Download function it is possible to obtain the information that Google holds about individual users for some of their products (Google, n.d.). However, as the data does not reside in the EU they are not required under EU Data Protection Law to detail exactly what data about a user they hold. Just from using the information obtainable, it is clear that Google has detailed knowledge of a person's email activity, the web searches they make, videos on YouTube that they have watched, calendar items and the location at which they have made searches. Google uses all this information about a person to serve targeted advertisements through their AdWords platform.

TRACKERS

Multiple types of trackers are available on the Internet today, they aim to find out every bit of information possible from a user while they visit a website. The most common form of tracking is used for Web Analytics, which primarily focuses on items such as "geo-location, company, all activity through time, first recorded entry, and session info. Full technical details, such as browser, device, platform, ISP, and operating system (OS)" (Opentraker, n.d.). Web Analytics is an impersonal and somewhat accepted form of tracking however; there are other trackers that follow a user's navigation through a website which are deemed to be much less acceptable but somewhat unavoidable with links being tracked by JavaScript. Additionally there is the ability to track a user's cursor movements as in such recent cases where "Facebook is

testing data mining methods that would silently follow users" mouse movements to see not only where we click, but even where we pause, where we hover and for how long" (Dr.Molle, 2012). Some people will see this extreme form of tracking to be very invasive.

The extent of these practices is currently expanding with every possible technique being explored, including browser profiling and tracking taking place. Browser profiling and tracking is the identification of an individual browser by its configuration and using this to track it. The identifying characteristics of the configuration held by the browser could be:

- User Agent,
- HTTP_ACCEPT Headers,
- Browser Plugin Details,
- Time Zone,
- Screen Size,
- Screen Colour Depth,
- System Fonts and
- if Cookies are Enabled (Panopticlick, n.d.).

Cookies are also used by 3rd parties to track users across websites, they use a unique ID to identify individual users and this is communicated to the 3rd party, whenever a user enters a website where the same 3rd party has a tracker active.

All forms of tracking mentioned, when tied into a user account of a website or service, can allow for data to be assigned to a specific individual. However this does not mean that if you are not signed into account that they will not attempt to track you. Facebook for example creates "Shadow Profiles" of non-Facebook users which uses cookies to track them following them from website to website with their 3rd Party tracking 'Like' buttons identifying non-users with characteristics previously mentioned (Charman-Anderson, 2011).

Privacy concerns are always highlighted by users throughout the topics in this section, and user's battle against the technologies tracking them with preventative solutions. An example of a well-known tracking blocker is Ghostery (Ghostery, 2014), which shows you the trackers on the website you are visiting and allows you to block them by default or as desired.

ADVERTISEMENTS

User tracking is used to enhance advertisements to adapt to a person's interests. This is called behavioural advertising that harvests data around a user's website visits, clicks and purchases to aim more appropriate adverts at a user. This enhances the chance that a user will click the advert to the paying website which may result in gaining new customers, increasing brand awareness and creating revenue streams. With behavioural adverts being a huge market, most websites adopt some form of advertisement and with privacy being questioned by people, preventative measures are once again available, either by blocking trackers previously mentioned or using Ad Blockers such as Adblock Plus (Adblockplus, 2014).

WHAT A COMPANY KNOWS ABOUT USER HABITS

A person's private online presence is not the only occurrence of data being freely released into other entities' hands. In the workplace, data is also being extracted from anything an employee does from web browsing to email and instant messaging. All of which can be monitored by a company through different means, such as:

- using Proxies to relay requests to the internet for an employee and therefore every website that is visited is logged
- It is possible to gain more information through more advanced forensics of data, e.g. Log data, cookies and cached data
- Packet sniffing is available to target browser and non-browser technologies such as instant messengers
- Companies can even go to the extreme of having surveillance monitoring software, which is able to monitor a devices screen making no action private.

This can lead to the company having an understanding of what an employee is doing at any given time. (Forensicon, n.d.)

MOBILE TRACKING

With approximately 64% of the population of the earth currently using a smartphone, this makes it a prevalent target for tracking the population around the globe (Emarketer, 2014). Smartphones and normal mobile phones all emit radio signals when turned on to ensure that they are registered onto a radio network. Utilising this functionality it is possible to track users across a large expanse if enough listening devices are used.

An example of this is a project called CreepyDOL which was developed by Brendon O'Connor of Malice Afterthought Inc. This utilises the WiFi signals emitted by SmartPhones. CreepyDOL is a distributed tracking system that uses low-cost hardware sensors to give near-real time identification of humans. The system uses Raspberry Pi's which listen for all WiFi packets that are sent by devices within its area, this information is then fed to a central location for analysis and identification on a map. Using the unique identifier that each device sends out it is then possible to track people as they travel across a city (O'Connor, 2013). This technology is not only being used by security researchers as hypothetical scenarios, it is also being used within shopping centres to track the footpaths of shoppers as they walk between shops

In a similar fashion to that of CreepyDOL, Path Intelligence have a product called FootPath which will monitor for 3G and 2G signals that any mobile phone will originate through an array of sensors within the shopping centre to a central processing area for analysis (Pathintelligence, 2014). Using this information Path Intelligence states to improve "Optimize Asset Performance, Boost Center Operations, Better your Marketing Return and Foster International Visitors".

A similar system was trailed using recycling bins within the City of London but was forced to be removed by the City of London Corporation (Warman, 2013).

PUBLIC CAMERAS

Cameras are watching our every move in public, our actions are data for the owners of the cameras to do with as they wish. With facial recognition becoming a common topic in today's news, and wearable devices fitted with cameras are being introduced to the market, examples of which are watches and glasses. Taking the example of Google Glass, there are many concerns about people with Glass recording others, with the extreme view that it will lead to facial recognition of the data Glass collects (Hill, 2014).

Wearable devices are not the only devices that are being used for facial recognition: the UK has been using CCTV footage to identify possible suspects in criminal activities (CBSNEWS, 2011). The use of cameras to identify subjects is becoming a common principle in major cities; London's Met Police have recently been fitted with body cameras (BBC, 2014). To what extreme public data is going to be used to find out information about passers-by it is unknown. As a result of the London riots a group people came together to work towards identifying people in pictures using facial recognition software. This data included data from Facebook accounts, information freely available or even provided to them by their own friends (Hill, 2011).

THE FUTURE

In the future you can expect to have less privacy than ever before in many circumstances, for instance: with technology just being created to have drones tracking your position by following your phone, a market to create smaller more advanced physical tracking units could be opened up (Goodin, 2014).

Companies are exploring different ways to bring your own devices (BYOD) into the office, however this creates a privacy challenge of what data is extracted and monitored by the company in the name of security, especially when the device is used for personal use. With a company potentially being able to tap into any information available on an employee's phone it is down to the user to be attentive to what a company can harness from their devices.

There is an ongoing fight on for privacy advocates, businesses and people in general. With technology developing and incidents arising such as governments intercepting more and more data about its citizens. This was demonstrated most recently by the UK government openly admitting it intercepts data going out of the UK to well-known services like Facebook and Google, raising the question as to what extreme an individual's data is compromised by government or other entities (BBC, 2014). Government interception of people's data could reveal any number of types of information from basics such as email addresses to supposedly private conversations.

Not knowing when any of your data is safe with any company due to malicious breaches or legal acquisition by governments opens up the question of: **are you giving away your data to a wider range of unknown entities when you trust websites?** As technology advances you can expect facilities to be available to track every moment of a person's life at a whim with little offered as a privacy shield.

CONCLUSION

The availability and volume of data about a user is dependent on how much they understand technology and care for their privacy. As mentioned, many people may not understand the extent to which information is gathered about them on the Internet and are oblivious to the behavioural advertising and profiling directed at them, of which they may not appreciate. On the other hand, there are users that generally do not mind giving away information and appreciate the possible benefits from data they provide to entities.

This returns us back to the two questions we have raised:

- How many people realise that information they post on social sites is also used by the owners of the website?
- Are you giving away your data to wider range of unknown entities, when you trust a website?

Privacy-conscious people on the other hand, have many options to prevent possible data loss on the Internet through the means of the items mentioned in this paper, such as Adblockers, Tracking Prevention and additional unmentioned ones such as VPNs, Tor and Proxies.

Due to the developing landscape of the Internet and the availability of data harvesting tools entities are using, privacy-conscious people must stay on top of new developments to ensure that they protect their personal data and stay ahead of the entities they do not wish to share their information with.

BIBLIOGRAPHY

Adblockplus, 2014. *Adblockplus Home Page.* [Online]
Available at: https://adblockplus.org/en/chrome
[Accessed 26 06 2014].
BBC, 2014. *Google and Facebook can be legally intercepted, says UK spy boss.* [Online]
Available at: http://www.bbc.co.uk/news/technology-27887639
[Accessed 26 06 2014].
BBC, 2014. *Metropolitan Police officers start wearing body cameras.* [Online]
Available at: http://www.bbc.co.uk/news/uk-england-london-27313500
[Accessed 26 06 2014].
CBSNEWS, 2011. *UK using facial recognition to hunt rioters.* [Online]
Available at: http://www.cbsnews.com/news/uk-using-facial-recognition-to-hunt-rioters/
[Accessed 26 06 2014].
Charman-Anderson, 2011. *Facebook finally admits to tracking non-users.* [Online]
Available at: http://tech.firstpost.com/news-analysis/facebook-finally-admits-to-tracking-non-users-208996.html
[Accessed 26 06 2014].
Dr.Molle, 2012. *Track ALL clicked elements using JavaScript.* [Online]
Available at: http://stackoverflow.com/questions/9509231/track-all-clicked-elements-using-javascript
[Accessed 26 06 2014].
Emarketer, 2014. *Smartphone Users Worldwide Will Total 1.75 Billion in 2014.*

[Online]
Available at: http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536
[Accessed 26 06 2014].
Forensicon, n.d. *Worker Beware – Employee Monitoring.* [Online]
Available at: https://www.forensicon.com/resources/articles/worker-beware-employee-monitoring/
[Accessed 26 06 2014].
Ghostery, 2014. *Ghostery Home Page.* [Online]
Available at: https://www.ghostery.com/en-GB/
[Accessed 26 06 2014].
Goodin, D., 2014. *Meet Snoopy: The DIY drone that tracks your devices just about anywhere.* [Online]
Available at: http://arstechnica.com/security/2014/03/meet-snoopy-the-diy-drone-that-tracks-your-devices-just-about-anywhere/
[Accessed 26 06 2014].
Google, n.d. *Google Takeout.* [Online]
Available at: https://www.google.com/settings/takeout
[Accessed 26 06 2014].
Hill, K., 2011. *'London Riots Facial Recognition' Vigilantes Abandon Their Project.* [Online]
Available at: http://www.forbes.com/sites/kashmirhill/2011/08/11/london-riots-facial-recognition-vigilantes-abandon-their-project/
[Accessed 26 06 2014].
Hill, K., 2014. *Google Glass Facial Recognition App Draws Senator Franken's Ire.* [Online]
Available at: http://www.forbes.com/sites/kashmirhill/2014/02/05/google-glass-facial-recognition-app-draws-senator-frankens-ire/
[Accessed 26 06 2014].
O'Connor, B., 2013. *CreepyDOL: Cheap, Distributed Stalking.* [Online]
Available at: https://media.blackhat.com/us-13/US-13-OConnor-CreepyDOL-Cheap-Distributed-Stalking-Slides.pdf
[Accessed 26 06 2014].
Opentraker, n.d. *Track Unique Visitors.* [Online]
Available at: http://www.opentracker.net/products/web-analytics/feature/track-unique-visitors
[Accessed 26 06 2014].
Panopticlick, n.d. *Panopticlick.* [Online]
Available at: https://panopticlick.eff.org/
[Accessed 26 06 2014].
Pathintelligence, 2014. *Pathintelligence Home Page.* [Online]
Available at: http://www.pathintelligence.com/
[Accessed 26 06 2014].
Solon, O., 2012. *How much data did Facebook have on one man? 1,200 pages of data in 57 categories.* [Online]
Available at: http://www.wired.co.uk/magazine/archive/2012/12/start/privacy-versus-facebook
[Accessed 26 06 2014].
Warman, M., 2013. *Bins that track mobiles banned by City of London Corporation.* [Online]
Available at: http://www.telegraph.co.uk/technology/news/10237811/Bins-that-track-mobiles-banned-by-City-of-London-Corporation.html
[Accessed 26 06 2014].

Sytel Reply is the Reply group company specialising in the Telecommunication, Media and Entertainment (TM&E) markets. The Sytel Reply mission is to support clients during their technology and business innovation processes by planning, developing and managing solutions for Networking, BSS and OSS and Mobile Applications within TM&E service provider market.Sytel Reply, thanks to its in-depth competence and experience, boasts a team of highly skilled professionals able to manage any end-to-end business and technology transformation programmes.

Sytel Reply
www.reply.com