Project Proposal

Introduction:
This project will investigate the large amounts of data given to businesses and the effect this might have on a malicious attacker's ability to steal user credentials. The project will focus on the social media site Facebook.

As we move into a new technological age thanks to the creation of the internet, companies can collect massive amounts of data on any individual and store it for long periods of time, making it a prime target for malicious users. This, coupled with users giving away data more and more freely means it could be only a matter of time before small pieces of information can be linked together to invade a user's privacy or worse open the user to attacks from harmful outside sources.

The recently implemented General Data Protection Regulation (GDPR) has curbed mass data acquisition by organisations with the "Right to be Forgotten" article which states "the data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay" (GDPR, 2018). With this regulation codified into law it is much harder for companies to keep user information once a user requests it be deleted, but onus is on the user to request the removal first which, for those not technologically savvy this leaves them unaware of this right. There is also part of the regulation that states pieces of personal data should be removed if they are "no longer necessary in relation to the purposes for which they were collected or otherwise processed"; this means data processed by a company will eventually have to be deleted in accordance with the regulation but is only useful for data that falls outside its original use.

The project is centred on cybersecurity which is "comprises technologies, processes and controls that are designed to protect systems, networks and data from cyber attacks" (Itgovernance.co.uk, 2018). The project will focus on helping increase understanding and develop techniques towards stopping "information-based de-anonymization" (Tian et al., 2018). It could also be used to make users more aware of how their information is used. Data can be used by organisations in many different ways, including:
Collecting a profile – Collating data from different sources to form a profile on the user
Targeted marketing – Tailoring products depending information gleaned from a user's online presence
Third party sales – Selling collected data to third parties to use for their own purposes
Malicious users gaining access to this data could use it to build their own portfolio on the user, following it up with an attack on their system or data. Making users aware of how both of these entities use personal data is imperative to keeping systems safe and keeping user's credentials secure.

Researching this data has become much more important with artificial intelligence and other data crunching technologies becoming more readily available. We may soon be at a stage where a user's credentials or other key information could be known using only seemingly unimportant data, without any significant details being known. The lack of privacy this encompasses is bad enough when these methods are employed by the organisations retaining the data but could soon also be a technology implemented by hackers or other malicious entities.

This increased data collection coupled with the new GDPR also has implications for the businesses working in this field. In America the cost of cybercrime is rated "at approximately $8.5 billion annually" (Romanosky, S. 2016). These company losses are expensive and with the GDPR stating data breaches can cost €20 million or up to 4% of annual revenue, whichever is higher, it has become imperative to protect against any credential loss.

Aims and Objectives:

Aim:
To assess the data an organisation collects, how it do so and ways in which this mass collection can create flaws for exploitation.

Objectives:

- Choose which social media platform will be used for the project
- Research data given to chosen social media when creating an account
- Research data given throughout the account's lifetime
- Research literature that discusses user data being leaked and the methods used to leak it
- Set up data capturing software to record for future tests
- Sign up to the chosen platform and give as little data as possible
- Sign up to the chosen platform and give as much data as possible
- Collate recorded data and analyse the results
- Repeat tests using different techniques and software
- Create final report and if appropriate use findings to eliminate weaknesses

Stretch Objective:

- Extend testing to different social media platforms or even other types of businesses
- Have volunteers create fake social media accounts for real time testing outside of controlled conditions
- Design a prototype social media that uses safer data practices found throughout the project
- Observe whether the centralised nature of internet business can be subverted to improve data privacy

Literature Review:

1. General Data Protection Regulation (GDPR, 2018)
   This regulation is now the standard for data protection in Europe and companies must abide by it. When researching organisations for this project the articles in this regulation will be important to understand how data moves around within a business.
2. Hype and heavy tails: A closer look at data breaches (Edwards et al., 2016)
   A journal article discussing Data breaches, how common they occur and how much damage they are expected to cause. It takes a sceptical view to other statistics that say the problem is getting massively bigger every year and tries to more accurately plot the data points. Because this article also states what types of attacks are used in some of its statistics, it will be easier to determine how data is usually stolen from organisations and narrow down the field of techniques used by malicious attackers.
3. Examining the costs and causes of cyber incidents (Romanosky, 2016)
   To be able to design a report to give cost effective solutions to an organisation's weak data security the costs to that organisation and its users must be known. This piece of literature calculates these costs that are placed on various industries affected by data breaches.
4. Deeply Understanding Structure-based Social Network De-anonymization (Tian et al., 2018)
   Because the project will focus on social media and data loss it's important to research methods already in use. This article discusses ways data that social media has attempted to anonymise can be rediscovered using de-anonymization. Because the researchers test the different algorithms used in the process their results may be used within the project.
5. Data breaches and identity theft (Roberds and Schreft, 2009)
   Social media businesses are not the only type of organisation that faces data breaches, payment networks also collect a lot of user data and so are very likely to be attacked. This journal article researches the monetary losses caused by these network having "too much data collection and too little security".

Project Plan:

To ensure that the project runs smoothly and is completed on time, a Gantt chart is shown below:

Gantt Chart

		Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13-17	Week 18	Week 19	Week 20	Week 21	Week 22	Week 23	Week 20	Week 21	Week 22	Week 23	Week 24	Week 25
		25-10-18	01-11-18	08-11-18	15-11-18	22-11-18	29-11-18	06-12-18	13-12-18	20-12-18	24-01-18	31-01-18	07-02-18	14-02-18	21-02-18	28-02-18	07-03-18	14-03-18	21-03-18	28-03-18	04-04-18	11-04-18
	Analysis/Research																					
1	Research data capturing software (Wireshark/Burpsuite etc)																					
2	Understand previous uses of scraped data in similar projects																					
3	Understand work environments used in similar projects																					
4	Review the terms and conditions of the company used																					
	Design																					
5	Setup and test several capturing parameters																					
6	Configure virtual work environment																					
7	Design a report to show recorded data																					
	Development																					
8	Use improvements in testing to create new reports																					
9	Complete Data collection and reports																					
10	Attempt Stretch objectives																					
	Contigency																					
11	Extra time to complete each task in case overun occurs																					

The chart shows the loop of researching new techniques to gather this data and then designing tests and virtual work environments to collect new data for the reports.

Below are the pieces of the plan put into more detail:

1 – Working to develop skills in areas of data capturing, this will probably involve researching software including Wireshark and Burpsuite.

2 – Using previous research completed by other researchers or users to gain new ways of using data capture software

3 – Using previous research completed by other researchers or users to gain new ways of setting up virtual environments for initial testing

4 – Within the companies legislation that a user signs to create an account details may be found on what data they collect and how they do so, a review of that legislation will be completed

5 – After researching methods of data capture and their associated environment, more rigorous testing will take place.

6 – Using initial tests and previous research to create a virtual working environment of which to do further testing out of

7 – Creating a report to display the data collected during these tests, this is an initial report

8 – After initial testing and with any improvements that come from it a new set of reports will be created to match new techniques

9 – A final collation of all data into reports and completion of the main task

10 – If work completes on time, some extra time is allocated to trying to complete the stretch objectives

11 – As stated in the risk analysis, a contingency for each stage is given to allow for overrun on parts of the project and keep on track

Risk analysis:

Risk Name	Risk Description	Severity	Probability	Mitigation
Active data collection	Penetration of a system not owned or under the control of any parties associated with the project	High	Low	Only commit passive observation and take no active interactions with any systems
Incorrect data Acquisition	Passive collection of user data not owned by any parties associated with the project	Low	High	As soon as it becomes apparent the data is not part of the project it is to be purged
Unwanted account interaction	Interactions online between project accounts and parties not associated with the project that are also using the social media service	Med	Med	Take steps so account is not discoverable to outside sources, failing that do not interact with these other users
Misunderstood intentions	Continuous testing on a network can start to look like a hacker probing for vulnerabilities, ending in system inquires or losing access to the system	Med	High	Space out tests as much as possible. Use networks that know the tests are being conducted
Insufficient backups	Keeping data stored on one system or keeping backups of the data in the same physical or virtual location, causing any data corruption to be absolute	High	Low	Several backups will be kept in several locations to make sure that if one is lost or damaged there are others to replace it
Participant Withdrawal	Participants may be used to create user accounts more realistically. They may wish to withdraw from the project at any time	Low	Med	They should be immediately allowed to do so and their data will be handled in accordance with the law
Missed Deadlines	For whatever reason, parts of the project take longer to complete than stated in the project plan	Med	Med	Contingencies have been added to the project plan in case this happens, giving extra time for many tasks

References:

Edwards, B. Hofmeyr, S. and Forrest, S.(2016) Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity. 2(1)* 3–14.

General Data Protection Regulation (GDPR). (2018). General Data Protection Regulation (GDPR) – Final text neatly arranged. Available at: https://gdpr-info.eu/ [Accessed 23 October 2018].

Itgovernance.co.uk. (2018). What is Cyber Security? | IT Governance UK. [online] Available at: https://www.itgovernance.co.uk/what-is-cybersecurity [Accessed 24 Oct. 2018].

Macwan, K. and Patel S (2018). k-NMF Anonymization in Social Network Data Publishing. *The Computer Journal.* 61(4) 601–613.

Roberds, W. and Schreft, S. (2009). Data breaches and identity theft. *Journal of Monetary Economics*, 56(7), pp.918-929

Romanosky, S. (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity.* 2(2) 121–135.

Selznick, L. and LaMacchia, C. (2018). Cybersecurity Liability: How Technically Savvy Can We Expect Small Business Owners to Be?. *Journal of Business & Technology Law. 13(2).*

Tian, W. Mao, J. Jiang, J. He, Z. Zhou, Z. and Liu, J. (2018). Deeply Understanding Structure-based Social Network De-anonymization. *Procedia Computer Science.* 129, 52-58.

.