2019

Facebook-Cambridge Analytica data harvesting: What you need to know

Ikhlaq ur Rehman
University of Kashmir, ak.edu05@gmail.com

Facebook-Cambridge Analytica data harvesting: What you need to know

Abstract

In 2018, it became public knowledge that millions of Facebook users' data had been harvested without their consent. At the heart of the issue was Cambridge Analytica (CA) which in partnership with Cambridge researcher, Aleksandr Kogan harvested data from millions of Facebook profiles. Kogan had developed an application called "thisisyourdigitallife" which featured a personality quiz and CA paid for people to take it. The app recorded results of each quiz, collected data from quiz taker's Facebook account such as personal information and Facebook activity (e.g., what content was "liked") as well as their Facebook friends which led to data harvesting of about 87 million Facebook profiles. The researcher then passed on this data to CA, which then used an algorithm that enabled them to psychologically profile people based on their Facebook interactions. Donald Trump had hired CA as a part of his 2016 Presidential election strategy. In order to deliver pro-Trump materials to individuals online, CA targeted individuals with a lesser known Facebook feature called "dark post" that contains personalized ads that are visible only to targeted individuals.CA consciously exploited fears of individuals with targeted advertising based on their personality profiles. The use of highly personalized ads made them vulnerable to Trump's messages that compelled them to vote for him. This digital onslaught played a significant role in Trump's victory over Hilary Clinton. This article will review how CA was able to harvest users' personal data and what were its repercussions.

Keywords

Cambridge Analytica, data harvesting, data leak, data privacy, Facebook data breach.

Introduction

Online Social Networks (OSNs) have turned out to be one of the most important means of communication for human beings in present times. OSNs such as Twitter, YouTube, Facebook, LinkedIn, Google+, to name as few, which have become one of the significant stages for social cooperation, such as, developing relationship, sharing individual experiences, and lending different services**.** Social network penetration worldwide is ever increasing. In 2017, 71% of the internet users were social network users which accounted for 2.46 billion users. In 2019, the estimated number will be around 2.76 billion **(Statista,2019)**. With the growth of OSNs privacy concerns have increased due to personal data shared online. Although, OSNs are deployed with privacy control mechanisms that protect user's private information from outside world. However, these privacy mechanisms can't still protect user's sensitive information from being getting leaked **(Li, Li, Yan and Deng, 2015)** The most widely used OSW is Facebook with 2.2 billion monthly

active users in July, 2018**(Statista,2018)**. Mark Zuckerberg, the co-founder and CEO advocates that Facebook's motive is to provide a communal service which clearly reflects on their homepage: "Facebook helps you connect and share with the people in your life". Facebook was founded in the year 2004 and since then it has grown into a billion-dollar industry with present worth of $541.5 Billion **(Forbes,2018)**. Pundits claim that Facebook takes advantage of its members and their social networks by exposing them to commercialization which raises concerns about users' right to privacy. They are of the opinion that Facebook privacy policies are often complex and lengthy and are mostly driven by profit motive **(Lilley, Grodzinsky and Gumbus, 2012)**. Most popular applications on Facebook Inc. have been allowing access to people's names, communicating identifying information, and in some cases, their friend's names to numerous advertising and internet tracking companies. This puts tens of millions of Facebook user's privacy under risk including the ones who set their profiles to Facebook's strictest privacy settings. This puts a question mark on Facebook's ability to keep identifiable information about their user's activities secure. Even though Facebook doesn't allow application makers transmitting data about users to outside companies even if a user agrees but some of its applications such as FarmVille, Texas HoldEm poker and FrontierVille have been found of transmitting user information to outside companies. Several applications that were communicating personal information were taken down. However, the actual reason for their unavailability still remains doubtful. **(Steel and Fowler, 2010)**. Recently the biggest Facebook data scandal came into the limelight where data from millions of users who used a popular personality application had been compromised by leaving it exposed online for anyone to access. The data for the personality quiz application myPersonality was distributed by academics at the University of Cambridge to hundreds of researchers via a website which lacked security measures, which led to it being left open to illegal access for four years **(Waterfield and Revel,2018)**.However, the most appalling thing was the fact that the firms had collected information not only of the users who had agreed to share their information with these services, but also anyone who was Facebook friends with those people. In early 2018 it became public knowledge that information of millions of Facebook users had been harvested by many firms including Cambridge Analytica **(Facebook Data Breach Essay,2018)** This work is an attempt to provide the insights of the Facebook data harvesting that is what data was harvested, how it was harvested, who harvested it and how was the harvested data was used.

myPersonality app

The myPersonality app was created by David Stillwell in 2007 at the Psychometrics Centre in the University of Cambridge. It motivated Facebook users to take psychometric tests up until 2012. During this time span data from 6 million volunteers was collected and shared with registered academic collaborators resulting in 45 scientific publications in peer-reviewed journals **(Kanter and Kanter,2018)**. Michael Konsinski joined Stillwell in 2008 and together they assessed Facebook users on five personality traits knows as the **Big Five**. **"**These are: openness (how open you are to new experiences?), conscientiousness (how much of a perfectionist are you?), extroversion (how sociable are you?), agreeableness (how considerate and cooperative you are?) and neuroticism (are you easily upset?). Based on these dimensions—they are also known as OCEAN, an acronym for openness, conscientiousness, extroversion, agreeableness, neuroticism". With all this information it is relatively easy to ascertain what kind of personality an individual is. After answering the questions users could opt-in to share their Facebook profile with the researchers. The method used by the researchers was very easy. The users were asked to fulfil a questionnaire in the form of an online quiz. From their responses the Big Five values of respondents were drawn by the psychologists and the compared with the online data form the respondents that is what they liked, posted or shared on Facebook, or what age, gender and place of residence they mentioned. Based on this, correlations were made by the researchers. It was astonishing to see how valid deductions could be drawn from online actions. For instance, fans of Lady Gaga were most likely to be extroverts, while those who "liked" philosophy were likely to be introverts and men who "liked" the cosmetics brand MAC were most probably to be gay. While it can be argued that information collected can't be an authentic way to produce reliable results but when tens, hundreds or thousands of data points are combined, the resulting predictions become remarkably accurate. In 2012, Konsinski and his team proved that it was possible to predict an individual's skin colour, their sexual orientation or their association with the Democratic or the Republic Party merely based on an average of 68 Facebook "likes" by a user. Even their intelligence, cigarette and drug use, alcohol and religious affiliations could also be determined **(Grassegger & Krogerus, 2017)**

Cambridge Analtyica (CA) – Its key Players

Cambridge Analytica is a British based political consulting firm which was set up in 2014 as a subsidiary of Strategic Communication Limited group(SGL). It was partly owned by Robert

Mercer-an American hedge fund billionaire. In 2014, Christopher Wylie, a Canadian data scientist went to work for Alexander Nix (former CEO, CA) at a company called SGL that specializes in influencing elections. Alexander Nix introduced Wylie to Steve Bannon (former Vice President, CA) who also happens to be the former White House chief strategist, as well as former head of Donald Trump's 2016 election campaign. They discussed their idea with Steve Bannon about how voter's opinions could be influenced during elections but for this project to work they needed money so Bannon introduced their idea to Robert Mercer. The idea was to combine micro-targeting which had existed in politics and then target individuals not just as a voter but also as a personality and eventually creating a psychological profile of each voter in a particular region or in this case the whole of the United States. Robert Mercer was impressed with the idea and invested $15 million in CA **(The Guardian,2018)**. The whole idea was based on 2014 paper by Cambridge University's Psychometrics Centre, "Computer-based personality judgments are more accurate than those made by humans"**.** To make the whole thing work CA needed data so Christopher Wylie first approached Michal Konsinski, one of the co-authors of the original myPersonality research paper in order to access the myPersonality database. But when negotiations failed, another psychologist and one of Konsinski's colleagues, Aleksandr Kogan offered them a solution that would replicate Stilwell's and Konsinski's original research **(Cadwalladr & Graham-Harrison,2018)**.

Cambridge Analytica- How 50 million Facebook records were harvested

Aleksandr Kogan developed an app called "thisisyourdigitallife" which featured a personality quiz. His company Global Science Research (GSR) in partnership with CA paid thousands of users to take a personality test and agreed to have their data collected for academic purposes. His application recorded the results of each quiz, harvested data from the taker's Facebook account such as "likes" and personal information and most importantly the data of their Facebook friends as well. The personality quiz results were then paired with their Facebook data such as "likes" to seek out psychological patterns. In order to be entitled for the personality test the user had to have a Facebook account and be a US voter so that tens of millions of the profiles could be matched to electoral rolls. From an initial trial of 1000 test takers, the researchers collected 160,000 profiles or about 160 per person. Within, a matter of months Kogan and Cambridge Analytica had a

database of millions of US voters which had its own algorithm to scan them, identifying possible political beliefs and personality characteristics. They could then decide who to target and craft their messages in a way that would possibly attract individuals and influence their actions or thoughts, also known a micro-targeting. The paid test- takers had agreed to share their data, and Facebook's default terms allowed their friends' data to be collected by an app, unless they had changed their privacy settings. But none of them had agreed to sell their data to firms such as Cambrdige Analytica that turned their data into a lucrative political tool **(Cadwalladr & Graham-Harrison,2018)**.

What happened to this data?

The harvested data from Facebook played a key role in Donald Trump's successful campaign against Hillary Clinton in the 2016 presidential election. The Trump campaign had hired Cambridge Analytica during 2016 election to run data operations, identify voters to target with ads, where to make campaign stops and help with strategic communication, like what to say in speeches **(Sherr, 2018)**. Instead, of persuading millions of voters across the United States to vote for Donald Trump the firm focused on tens of thousands that they knew through their analysis were hesitating. They defined 32 types of personalities across the country based on the information they had on electorate. They targeted individuals that were believed to be worried or neurotic with highly personalized messages, thus vulnerable to Trump's messages. The firm identified many such individuals/voters in three states: Michigan, Wisconsin and Pennsylvania which could swing the results in Trump's favor. In order to reach these targeted individuals, they used a very less known Facebook feature called "Dark Post". These are personalized ads on social media that are visible only by the person who is specifically targeted. In this case if the firm knew an individual was in favor of carrying fire arms then he would be targeted with the following message "Did you know Hillary Clinton wants to take your gun away?". Based on his habits and digital finger prints this ad would be sent to his Facebook news feed at a specific time that could be seen only by him and eventually disappearing in a few hours. The idea was to target voters with negative messages against the rival candidate in the election. This digital onslaught helped Trump gain 77000 votes in these three key states which eventually carried him to the victory **(Camelot,2018)**.

How did Facebook react?

The way Kogan had gained access to millions of Facebook profiles in 2013 made Facebook change its platform in 2014 to curb the data apps could access. This meant apps could no longer access your friend's data unless authorized by them. Developers also needed validation from Facebook before they could request any sensitive data from people. In 2015, Facebook from the journalists at The Guardian learned that Kogan had shared the data of users with Cambridge Analytica without their consent and therefore breaching its policies. So it banned Kogan's app and demanded Kogan and Cambridge Analytica to destroy all the data that they had gathered improperly. Facebook said it received certifications they had complied. However, in 2018 Facebook learned from The New York Times, The Guardian and Channel 4 that Cambridge Analytica might not have destroyed the data as they had certified so it banned Kogan and Cambridge Analytica from their platform. Mark Zuckerberg said that "This was a breach of trust between Kogan, Cambridge Analytica and Facebook. But it was also a breach of trust between Facebook and the people who share their data with us and expect us to protect it. We need to fix that". Facebook had already taken some steps in 2014 to prevent data apps harvesting people's information in the way Kogan's app did. However, to make its platform more secure Mark Zuckerberg highlighted some new measures that Facebook is going to take. First, "we will investigate all apps that had access to large amounts of information" before Facebook changed its platform in 2014 to curb the data apps could access. Further apps will be audited for any suspicious activity and app developers will be banned from the platform if they don't adhere to a full audit. If Facebook finds out that users' data has been misused, then anyone who has been affected will be notified. Second, Facebook will cut down the problem form its source by curbing the data apps and developers can access. Facebook "will restrict developers' data access even further to prevent other kinds of abuse". For instance, developers' access to data will be removed if a user hasn't used their app in three months. Further, while signing in an app's access to data will be limited to users' name, profile photo and email address. Developers would not only need validation from Facebook but also sign in an agreement before they can ask users for access to their posts and other sensitive data. The third and final step will enable users to manage and understand which apps have access to their data. Facebook will provide a tool "at the top of your News Feed" that will show users the apps they have used "and

an easy way to revoke those apps' permissions" to their data. Although, this tool was already available in privacy settings, but to make sure everyone sees it Facebook put it at the top of news feed **(Zuckerberg,2018)**.

Conclusion

The Facebook data privacy scandal led to the harvesting of millions of Facebook profiles by Cambridge Analytica. The firm was able to get access to private information of Facebook users due to number of factors, mainly including insufficient safeguards against data mining firms, inadequate supervision of developers by Facebook and users excessively agreeing to Facebook terms and conditions. In 2014 Aleksandr Kogan had developed a Facebook app called "thisisyourdigitallife" which allowed people to participate in a personality quiz test and Cambridge Analytica paid for people to take it. This app not only harvested data from the quiz taker's Facebook account but also the data of their Facebook friends. This led to data harvesting of up to 87 million Facebook profiles **(Kang and Frenkel, 2018)**. Kogan, then shared this data with Cambridge Analytica, a political consulting firm that uses data to determine voter personality traits and behavior. This led to violation of Facebook's terms and conditions which forbids the sharing or sale of data "to any ad network, data broker or other advertising or monetization-related service" **(Granville,2018)**. Facebook learned about this in 2015 and it immediately banned Kogan's app from its platform and demanded Kogan and CA to delete all the data that they had gathered illegally. Kogan and CA all certified to Facebook that they deleted the data. However, in March 2018 Facebook learned that the data wasn't deleted after CA whistleblower Christopher Wylie revealed that the data harvested form Kogan's app was used for building "psychographic" profiles of people and deliver pro-Trump material to them online **(Meredith,2018)**.CA used the data for clear political purposes- to help conservative campaigns in the 2016 election, including Donald Trump's campaign **(Rosenberg,Confessore and Cadwalladr,2018)**. Facebook first responded on March 17, 2018 in a Facebook post by Paul Grewal, VP & Deputy General Counsel, who wrote that, "The claim that this is a data breach is completely false. Aleksandr Kogan requested and gained access to information from users who chose to sign up to his app, and everyone involved gave their consent. People knowingly provided their information, no systems were infiltrated, and no passwords or sensitive pieces of information were stolen or hacked"**(Grewal,2018)**.On the

same day Alex Stamos, Facebook's Chief Security Officer, tweeted (later deleted the tweet) that, "Kogan [one of Cambridge Analytica's researchers] did not break into any systems, bypass any technical controls, or use a flaw in our software to gather more data than allowed. He did, however, misuse that data after he gathered it, but that does not retroactively make it a 'breach.'"**(Johnston,2018)**.International Organization for Standardization and the International Electrotechnical Commission – two bodies that regulate global security practices, define data breach as follows "a compromise of security that leads to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to protected data transmitted, stored or otherwise processed"**(ISO,2015)**.Since the Facebook systems weren't bypassed and the data was misused by a third party that clearly violated Facebook terms and conditions, the incident therefore doesn't qualify as a data breach as understood by the global cyber security community. Consumers must realize that their data are worthwhile. Consumers should learn how companies, in particular those offering free services such as Facebook and Google, use their personal data to run their business. Consumers should read data privacy notices and make use of the in- product user controls offered by most tech companies. Consumers should take advantage of their rights to ask a company to have their personal data viewed, edited and deleted because, after all, data belongs to consumers and not companies. If companies engage in illicit or false data handling practices, consumers should file complaints with the Federal Trade Commission (FTC) or any other governing bodies. Lastly, consumers should promote greater transparency and company controls and require their elected officials to do more to protect privacy. Almost, all the companies in the world now process personal data electronically. So, companies ought to learn to better balance privacy risks with privacy controls. The riskier the use of data, the more user controls are necessary. Controls can include user friendly and distinguished privacy notices, clear consent and privacy-friendly default settings. While sharing the data from third parties, companies should make sure that those companies comply with their privacy standards by investing in period audits. Similarly, when receiving data from third parties, companies should make sure that the data was collected a proper manner, not merely believing in vendor's word, but again, by performing period audits. And eventually governments must reshape outdated laws in order to tackle the current complexities of data usage and transfers The European Union, for instance has set up a global example, through the General Data Protection Regulation that came into effect on May 25, 2018. "This is a comprehensive piece of legislation that (1) expands data subjects' rights (2) enforces 72-hour data

breach notifications (3) expands accountability measures and (4) improves enforcement capabilities through levying fines of up to 4% of global revenue." This law is applicable only to European countries, but most multi national tech companies have incorporated these standards for all their customers**(Kozlowska,2018)**. However, time has come that similar privacy protection laws are passed across the globe so that everyone can benefit from the opportunities that the 21st digital economy brings with it.

References

Cadwalladr, C., & Graham-Harrison, E. (2018). The Cambridge Analytica Files. *I made Steve Bannon's psyhological warfare tool': meet the data war whistleblower*. Retrieved from https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). How Cambridge Analytica turned Facebook 'likes' into a lucrative political tool. Retrieved from https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm

Camelot, S. (2018, February 16). *Your Facebook data used for Trump & Brexit win by Cambridge Analytica* [Video file]. Retrieved from https://youtu.be/q6wFDhUx1NM

Facebook Data Breach Essay (Free Example). (2018, December 19). Retrieved from https://www.aceyourpaper.com/essay/facebook-data-breach-essay/#Related_Topics

Forbes (2018). Global 2000: Growth Champions. Retrieved from https://www.forbes.com/companies/facebook/#79fa3c1e4193

Granville, K. (2018, March 19). Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. Retrieved from https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html

Grewal, P. (2018, March 17). Suspending Cambridge Analytica and SCL Group from Facebook. Retrieved from https://newsroom.fb.com/news/2018/03/suspending-cambridge-analytica/

ISO/IEC 27040:2015. International Organization for Standardization. Retrieved from https://www.iso.org/obp/ui/#iso:std:iso-iec:27040:ed-1:v1:en

Johnston, C. (2018, March 19). Cambridge Analytica's leak shouldn't surprise you, but it should scare you. Retrieved from https://theoutline.com/post/3796/cambridge-analyticas-leak-shouldnt-surprise-you-but-it-should-scare-you?zd=2&zi=ru55e42f

Kang, C., & Frenkel, S. (2018, April 4). Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users. Retrieved from https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html

Kanter, J., & Kanter, J. (2018, May 15). Facebook is investigating another app created by Cambridge University academics after it hoovered up the data of millions of users. Retrieved from

https://www.businessinsider.in/facebook-is-investigating-another-app-created-by-cambridge-university-academics-after-it-hoovered-up-the-data-of-millions-of-users/articleshow/64173953.cms

KOZLOWSKA, I. (2018, April 30). Facebook and Data Privacy in the Age of Cambridge Analytica. Retrieved from https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/#_ftnref19

Krogerus, M., & Grassegger, H (2017, January 28). The Data That Turned the World Upside Down. Retrieved from https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win

Li, Y., Li, Y., Yan, Q., & Deng, R. H. (2015). Privacy leakage analysis in online social networks. *Computers & Security*, *49*, 239-254.doi: 10.1016/j.cose.2014.10.012

Lilley, S., Grodzinsky, F. S., & Gumbus, A. (2012). Revealing the commercialized and compliant Facebook user. *Journal of information, communication and ethics in society*, *10*(2), 82-92.doi:10.1108/14779961211226994

Meredith, S. (2018, March 21). Here's everything you need to know about the Cambridge Analytica scandal. Retrieved from https://www.cnbc.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html

Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018, March 17). How Trump Consultants Exploited the Facebook Data of Millions. Retrieved from https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Sherr, I. (2018, April 19). Facebook, Cambridge Analytica, data mining and Trump: What you need to know. Retrieved from https://www.cnet.com/news/facebook-cambridge-analytica-data-mining-and-trump-what-you-need-to-know/

Statista. (2018). Most popular social networks worldwide as of October 2018, ranked by number of active users (in millions) Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Statista. (2019). Number of social media users worldwide 2010-2021 Retrieved from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Steel, E., & Fowler, G. (2010). Facebook in privacy breach. *The Wall Street Journal*, *18*(1). Retrieved from https://www.wsj.com/

The Guardian. (2018, March 17). *Cambridge Analytica whistle-blower: 'We spent $1m harvesting millions of Facebook profiles*'. [Video file]. Retrieved from https://youtu.be/FXdYSQ6nu-M

Waterfield, P., & Revell, T. (2018). Huge new Facebook data leak exposed intimate details of 3m users. Retrieved from https://www.newscientist.com/article/2168713-huge-new-facebook-data-leak-exposed-intimate-details-of-3m-users/

Zuckerberg, M. [Mark]. (2018, March 21). I want to share an update on the Cambridge Analytica situation -- including the steps we've already taken and our next steps to address this important issue. We have a responsibility to protect your data, and if we can't then we [Facebook status update]. Retrieved from https://www.facebook.com/zuck/posts/10104712037900071?pnref=story