2019

Reclaiming Control over Personal Data with Blockchain Technology: An Exploratory Study

Thomas Mejtoft

David Hellman

Ulrik Söderström

32ND BLED ECONFERENCE
HUMANIZING TECHNOLOGY FOR A SUSTAINABLE SOCIETY
JUNE 16 – 19, 2019, BLED, SLOVENIA, CONFERENCE PROCEEDINGS
A. Pucihar (ed. et al.)

University of Maribor Press

Reclaiming Control over Personal Data with Blockchain Technology: An Exploratory Study

THOMAS MEJTOFT, DAVID HELLMAN & ULRIK SÖDERSTRÖM

Abstract With the digitalization and increasing number of Internet users, more and more personal data breaches occur. Many people are not aware of their personal data rights and have not received any instructions on how to act in situations such as when their personal data is abused. This is something that illustrates the flaws of the Internet. A technology that provides solutions to some of these problems, such as trust and transparency, is the blockchain technology. Hence, the objective of this paper is to investigate knowledge about personal data rights and to explore the design of a prototype of a blockchain application for increased security and transparency. User tests were conducted, highlighting the greatest needs for users to feel secure and in control over their personal data. This knowledge provide the foundation for a prototype based on blockchain technology that gives the users increased security and forces those who store personal data to be more transparent with the usage.

Keywords: • Blockchain • Personal data • User test • Exploratory Study • Internet •

CORRESPONDENCE ADDRESS: Thomas Mejtoft, PhD, Associate Professor, Umeå University, Department of Applied Physics and Electronics, Digital Media Lab, Umeå, Sweden, e-mail: thomas.mejtoft@umu.se. David Hellman, Umeå University, Department of Applied Physics and Electronics, Digital Media Lab, Umeå, Sweden. Ulrik Söderström, PhD, Associate Professor, Umeå University, Department of Applied Physics and Electronics, Digital Media Lab, Umeå, Sweden, e-mail: ulrik.soderstrom@umu.se.

DOI https://doi.org/10.18690/978-961-286-280-0.22 ISBN 978-961-286-280-0
Dostopno na: http://press.um.si

1 Introduction

The Internet has been around since the late 1960s (as Arpanet) and has both seen a steady user growth (Leiner et al., 1997; Statista, 2018) and become a natural part of everyday life. While originally intended as a medium for peer-to-peer communication, the development of the World Wide Web, in the early 1990s, provided an easy to use platform with both high value and a significant reach and usage (Leiner et al., 1997). When the Internet was invented, the focus was to solve communication. Thus, it was not built to be a system without centralized control (Sweeney, 2015; Rainie & Anderson, 2017). This has meant that the responsibility for online security and privacy have been in the hands of the users and those providing content.

Much of the content on the World Wide Web can be accessed easily and for free. However, when using a service on the Internet, the users often leave traces that are stored in the systems, so-called digital footprints (Internet Society, 2018). This can be voluntary, by e.g. sharing personal data in a registration process, or involuntary, through e.g. the collecting of cookies that track the users' movements (Internet Society, 2018). What many users are not aware of, however, is what happens to their personal data and how companies profit from it.

There is an element of risk involved when using the Internet and users themselves have to assess whether this risk is critical or not. The perceived risk may not be easy to determine, which is why a certain amount of trust is necessary for Internet usage (Cerf, 2010). As the recently highlighted case with Cambridge Analytica show, abuse of data can have serious implications (Rosenberg, Confessore & Cadwalladrmarch, 2018). New approaches for users to feel safer and more secure on the Internet are required in order to maximize the value of the services provided on the Internet. Thus, increased transparency on how personal data is being used has to improve. There is no central authority running the Internet, therefore, there is no single person or agency regulating content abuse or personal data breaches. A solution to this could be an authority acting as a trusted third party (middleman). The middleman could then make sure that both parties uphold their bargain of not abusing personal data nor the provided service. This would be feasible but not optimal as the middleman could lose all data that they have been entrusted with in a hacker attack or the middleman could be a malicious actor. From this example, parallels can be drawn to economic

transactions. Similarly, for transferring money, there is a counterparty risk, i.e. a risk that the other part will not fulfil their obligations. Today, national banks usually act as the middleman between the trading peers when we discuss transactions involving money. In exchange for removing (or at least lowering) the counterparty risk, the middleman often charges fees and dictate the rules of the transactions.

For the purpose of removing counterparty risk without having a middleman, blockchain technology can be applied and replaces the trusted third party in the transaction. It is a technology that enables instant and secure peer-to-peer transactions (Nakamoto, 2008). The blockchain has several use cases. For example, increasing productivity in supply chain management, voting and, maybe the most pronounced, economic transactions. One of the most commonly known applications to be built out of the blockchain technology is the cryptocurrency Bitcoin. Ideas of using blockchains to protect personal data and increase privacy have both been deemed important and discussed in research (Zyskind, Nathan & Pentland, 2015).

The objective of this exploratory study is to discuss how to increase personal data security through blockchain technology. This will be achieved in two steps: First, by mapping the current knowledge and perception of online personal data management. Second, from the gathered data, presenting a prototype using blockchain technology for increasing personal data security and usage transparency.

2 Theoretical Framework

In 2008, a whitepaper was released in a cryptographic community by an unknown author using the pseudonym Satoshi Nakamoto (Norman, 2017; Nakamoto, 2008). The paper proposed a solution for a decentralized electronic cash system called Bitcoin. The paper introduced the blockchain technology by proposing solutions to some complex cryptographic problems. Back in the 1990s attempts to create cryptocurrencies were made but without any significant impact. However, some of these ideas laid the foundations for this description of Bitcoin. The ripples created by the 2008 whitepaper have had wide spread impact since its release. While Bitcoin and other currencies continue to develop, blockchain technology has been used to create many new applications where trust is needed

among people that do not know each other. This can be e.g. to counteract diamond theft, streamlining stock markets and verifying contracts (The Economist, 2015; Iansiti & Lakhani, 2017).

2.1 The Double Spend Problem

Since trading with cryptocurrencies does not involve any physical exchange, there is a risk of a token being spent in multiple transactions which is called the double spend problem. The problem has been one of the biggest obstacles in realizing the use of cryptocurrencies. The blockchain technology solves this problem with a combination of a distributed ledger, a public record of transactions which works as a proof of information, and adding a level of complexity for creating a new block.

Nodes are central in the blockchain technology and can be any Internet connected device. All nodes on the network keep an updated version of the same ledger. If an alteration of a previous transaction is made by one node, the blockchain would become forked because the two versions of the ledger should be identical. To prevent fraudulent branches, the longest chain is considered true. Therefore, the creation of a block was made time-consuming through hash functions, making it difficult to grow the false branch faster than the true one. For a node to create a new block, a specific value has to be found. All active nodes on the network compete to find this value in the commonly used Proof of Work method. Thus, an attacker would need to outpace the other nodes in order to create a longer chain from a new branch. This process makes it very unlikely that attacks would succeed (Nakamoto, 2008).

2.2 Incentives

A blockchain would not exist without the creation of new blocks and honest participants. Therefore, incentives are necessary for attracting participants, keeping them honest and making them create new blocks. There are several different ways to incentivize honest participants, the two most popular ways are Proof of Work and Proof of Stake, and both reward creation of a new block.

Proof of Work is essential to blockchains such as Bitcoin. It is a piece of data that is very time consuming to produce but at the same time very easy to verify if it is correct once created. In order to keep the blockchain reliable, the longest chain, i.e. with the most proof of work in it, is considered the honest chain. When creating a new block, all mining nodes on the network compete to find a specific value. Once the value is found and the block is created, the creator is rewarded with a token (e.g. crypto currency). In order to prevent improved hardware to allow faster creation of blocks over time, the difficulty of finding the value for the block is increased in correlation to creation speeds, i.e. numbers of miners (Nakamoto, 2008; Norman, 2017). Since Proof of Work is based on computing power, it is also based on the idea of using as much energy as possible as fast as possible. This is one of the major critiques against blockchains. Proof of Stake is alternative to Proof of Work and, hence, another method of giving incentives to the creators of honest nodes. Compared to Proof of Work , the nodes on the network do not compete for creating a new block. Instead, the creator of a new block is chosen from a pool of users based on their percentage of total tokens (Norman, 2017). In other words, the creators have something at stake in the system. The creator of a block is rewarded with transaction fees for the transactions that go into the block. Because the nodes are not competing for creating new blocks with the Proof of Stake method, the energy consumption required is a lot less compared to that of Proof of Work.

2.3 Digital Footprint

Digital footprints are the traces that users leave when using the Internet (Girardin, Blat & Ratti, 2008). This could be online behavior, email records or personal data. One of the problems with digital footprints is described by Internet Society (2018) as: "Every day, whether we want to or not, most of us contribute to a growing portrait of who we are online; a portrait that is probably more public than we assume". These digital footprints mean companies can gather large amounts of data about their users without their users knowing. The data sets are of commercial value and thus can be sold. It is therefore not uncommon that personal data is sold without the consent of the users. Studies conducted on British citizens made by Information Commissioner's Office in 2014 (ComRes, 2014) show that there are deficiencies in the awareness regarding personal data protection of individuals. Furthermore, more than half of the

participants were concerned about organizations collecting and holding information online.

As been mentioned earlier, individuals disclose an increasing amount of information about themselves. Both in terms of data collected and in terms of sharing and user generated content (Blackshaw & Nazzaro, 2006; OECD, 2007). Therefore, privacy has certainly become a hot topic in current times. Privacy has been defined by Warren & Brandeis (1890, p. 193) as "the right to be let alone" and today privacy has been adressed mostly in terms of the choice that we make to share our information to get better services and be part of social communities as well as privacy risk (Appelgren, Leckner & Mejtoft, 2014; Acquisti, Taylor & Wagman, 2016).

2.4 Data Protection Regulation

In 2018 a new legislation called General Data Protection Regulation (GDPR) was implemented within the EU (European Union, 2016; Datainspektionen, n.d.). The legislation is a stricter version of the earlier Data Protection Act. GDPR have an impact on many companies worldwide since it is not only covering companies operating in the European Union (EU), but also companies that handle information of EU residents. The major impact of the new legislation is:

- Penalties will be more severe for misdemeanor.
- Consents must be clear and accessible.
- Rights to get personal data erased.
- Transferring data to third parties should be informed to the users.
- Increased security requirements.

3 Methodology

In order to design a blockchain solution that addresses the users' need for data security, the research was divided into two parts. An initial survey was undertaken and this was complemented by a qualitative research on a constructed prototype. These experiments were conducted during spring 2018.

3.1 Participants

An initial survey with 12 people was conducted. Since the Internet is widely used (Statista, 2018), a broad age range of participants, both in terms of age and gender, was considered important for the study. For ethical considerations, no one below the age of 18 was included. After analyzing the results of the survey, a prototype was built and tested. The users were presented with a set of instructions, thereafter they had to navigate through a hi-fi prototype, created in the digital design tool Sketch.

The survey questions were divided into two parts. The first consisted of seven questions regarding concerns, online behavior and data rights. The second part consisted of questions regarding valuation of personal data and customization. Some of the questions asked were: "Does a company that lose a customer's personal data have any responsibility to the customer?", "You are informed that a company, where you recently registered, has abused your personal data. What do you do?" and "If the quality of a service would improve with your personal data, would you be willing to share it? E.g. products with nuts would not be shown if you are allergic."

3.2 Prototype Testing

With the knowledge gathered from the desktop research of the blockchain technology and the survey data, a hi-fi prototype was created in Sketch. The purpose of testing the prototype was to investigate the participants' perception of what service the application provided and whether they felt safe using it. The method chosen was structured interview (Fontana & Frey, 2005). A scenario was presented to the participants who were supposed to navigate through an interface while the interviewer noted comments and the navigation patterns of the respondents. After the test was finished, the participants were required to answer a few questions.

> *The scenario: The user was looking for a new pair of running shoes. After searching, a suitable pair were found, although, from a previously never visited website. The user decided to buy the shoes, verifying and sharing his or her personal data with a blockchain application. When coming online, later on, the user was faced with a personal data access request from a suspicious website (Figure 1). The user could see*

what data was being requested and where the website got it from the first place. The user then made a choice of whether to accept or decline the request.

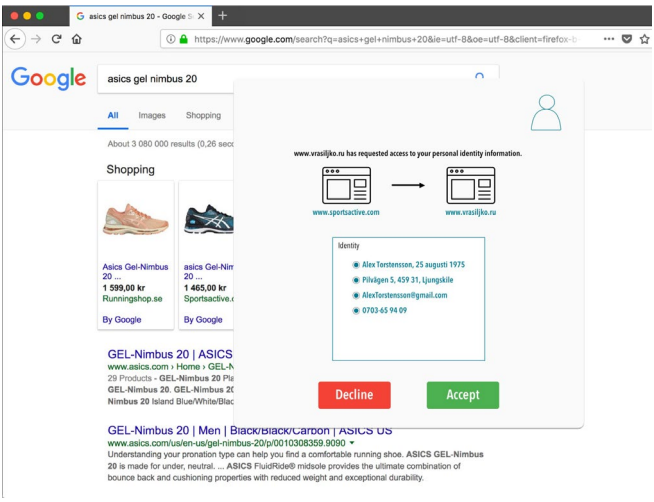

Figure 1. Access request from suspicious website.

4 Results

The results of the survey indicate that people may act very differently when there either is suspicion or confirmation of personal data abuse. In one of the questions in the survey, the respondents had to specify what level of control they perceived to possess over their personal data online. The overall responses were slightly pessimistic with most answers being neither in full control nor no control. However, the second most common answer was close to no control.

This data was collected just before the GDPR was introduced and under GDPR a company that loses a customer's personal data has certain responsibilities, including compensating for the losses. Despite being one of the alternatives in the questionnaire none of the respondents selected this answer. In fact, the responses showed a wide variance in the level of knowledge about the responsibilities of companies in the case of a data breach.

The results from the survey demonstrated that many of the respondents did not know what cause of action was open to them if their personal data was abused. This is shown by the answers to the question on what they would do if they were informed that a company abused their personal data, the answers differed a lot.

A blockchain personal data application would provide the users with a possibility to sell their personal data. However, a majority of the respondents stated that they were not willing to sell any data even if the option was available to them.

4.1 Personal Data Blockchain Application

One of the objectives with this project was to present a proposal of a blockchain application for personal data security. This was during the project prototyped based on the gathered data and tested.

From the survey, a number of key components for the prototyped blockchain application for personal data were identified. The proposed solution provides control and responsibility for personal data to its users. As a decentralized application, it exists in a layer on top of a blockchain network. Hence, the application does not need any own monetary cryptocurrency but instead use a utility token. Every user will have a personal token of their personal data. This token will be referred to as Personal Identification Token or PIT for short. For sharing single pieces of personal data, this token can be split up into minor parts. All personal data is stored locally in a token, only accessible after consent through biometric identification. Thus, every new usage of a PIT has to be confirmed by the user.

Whenever a company requests any personal data, a smart contract is enabled and the transaction is verified by the blockchain network. The smart contract will persist, which means any violations of the agreement would see the transaction revoked.

The application also offers alternative ways of identification. For example, in an online purchase, a user's identity can be verified by the network without having to be shared with the vendor. Thus, the only data that has to be provided to the vendor is the shipping address. The network verification also provides the vendor with information that the customer, in fact, is authentic. In a similar way,

users could, for example, confirm that they are above a certain age in a reliable manner, without the users having to provide their social security number.

Verification. The need for valid identification will still persist with the presented solution. Even if the network removes the need for a third party identification guarantor, a trustworthy authority still has to issue the identifications. This means reliable authorities are essential to the application. Before adding any personal data to the user's token, unique and critical data such as social security number has to be verified by these authorities.

Incentives. In order to make transactions possible, new blocks have to be created. For keeping the nodes on the network honest and incentivizing block creation, Proof of Stake will be used. The block creators will be rewarded with transaction fees for the transactions stored in their block. Because the personal data blockchain does not use any monetary tokens, the transaction fees collected will be in the cryptocurrency of the underlying blockchain. The chance to be chosen to be the creator of the next block will be equal to all participants on the network. Using Proof of Stake also give the system a more sustainable approach.

Transparency. The blockchain technology allows the application to display usage of personal data more clearly, i.e. to make usage more transparent. Should a company try to sell personal data to a third party, the transaction would have to be verified by the user before the other party could access any personal data from the PIT. Furthermore, who actually transferred the PIT to the third party would be visible for the users. This increases the demand on companies to respect their customers' personal data.

The application offers an overview of whom have access to the personal data through smart contract agreements. This provides the possibility to revoke access to personal data. In terms of the smart contract, the agreement is broken if data is revoked. Thus, smart contracts should be designed to protect both parties in the case of a sudden revoke of a transaction, meaning one party should not be able to profit on behalf of the other.

Security. Local storage of personal data means the security responsibility lies in the hand of the users. Thus, users verify transactions of personal data with biometrical methods available in contemporary smart devices, such as a

fingerprint scanning. This provides security against possible attacks. Unless the attacker manages to hack the biometric identification, the data will not be accessible.

Because the blockchain technology solves the problem of double spending, PIT transaction abuse is unlikely. A fraudulent company trying to distribute a PIT without consent will have to outpace the rest of the network. Because biometric identification is needed to unlock the PIT, fraudulent transactions would not be viable.

4.2 Prototype Testing

Six of the respondents from the initial survey were chosen to participate in the testing of the hi-fi prototype. A pilot test was run on one of the participants to test the prototype. The initial steps for the user to access the prototype were shown to be somewhat abundant, hence, they were removed for the other tests.

Following the suggestions of Fontana and Frey (2005), the participants were presented only with the necessary instructions. Those who encountered minor navigational confusions had to solve the problem without any support from the interviewer. After navigating through the interface, the participants were asked to describe what had happened during the navigation. This provided indications that the participants had the desired perception of what occurred during the test. The last step of the test featured a PIT usage request from a suspicious website. The users were also presented with information from what source the suspicious website actually had acquired the PIT. However, this information seemed to be bypassed with three participants, who did not register the information and took no action.

5 Discussion

The results from the user tests showed that individual's knowledge of the Internet differs significantly. Thus, applications such as the one proposed in the prototype are needed to assist people by making information accessible or subconsciously understood. The Internet was not designed to always protect the user, rather it is merely a means of communication (Sweeney, 2015; Rainie & Anderson, 2017). The decentralize nature of the Internet together with the

current usage of Web 2.0 platforms such as social media mean trust is an essential component (Cerf, 2010). Even though the solution to trust does not necessarily need to be one using blockchain technology, this research illustrates that it can be used for addressing problems with online trust. Thus, a blockchain solution is of high relevance to all Internet users.

This study displayed several flaws in knowledge about what rights people possess regarding personal data. The study was performed around the time of the implementation of the GDPR directive (European Union, 2016) within the EU and the results do not show the effects of the GDPR. The old legislation, the Data Protection Act, can be said to have failed in some respects. It has not provided enough protection and many people are unaware of their rights, probably due to a lack of information around personal data security (ComRes, 2014). Because of the digitalization of business functions and the steady Internet user growth (Statista, 2018), the need for an updated version of personal data legislation has been required for quite some time. Recent developments with Cambridge Analytica (Rosenberg et al., 2018) influence in the EU referendum and the US presidential election highlight the need even further. The harvesting and usage of personal data can have major implications, something both companies and users of the Internet have to become aware of.

While a majority of the survey respondents stated that they were not willing to sell any personal data, several of them would be willing to share data for free in order to get a more customized online experience. Either the questions were unclear, making the respondents confused or the notion of selling the personal data was intimidating. Data is of great value to companies, while something that is quite intangible for users. Consider, for example, that the data sold by Facebook to Cambridge Analytica (Rosenberg et al., 2018) that, alleged, had an impact on the outcome of the US Presidential Elections in 2016. A significant amount of money goes into these campaigns for the purpose of marketing and buying data to target campaigns.

Even though the survey conducted was carried out a few months after the Cambridge Analytica case, the respondents rated their perceived control over their personal data to be high. Combining this knowledge with the widespread of answers on how to handle personal data abuse indicates that the respondents

are generally secure in their handling of personal data or that they are not aware that the data is actually being abused without their consent.

The proposed Blockchain solution would limit the possibilities for companies to make money out of the users' data. It could even make them have to pay to access personal data. Furthermore, it would provide the verifiers with more money and potentially the users too. Hence, its implementation would probably encounter a lot of resistance. As with all technology, especially one that has not been thoroughly researched, there is a risk that technological advancements disrupt the application. Examples of this could be possibilities to fake biometric identification. By keeping the network participants incentivized, they will likely stay honest and the blockchain network will remain strong.
This exploratory study provides some interesting results, but to further strengthen the results and develop a useful application, more studies are needed with a wider user base.

6 Conclusion

What the effects of the GDPR legislation will be are yet to be fully observed. However, as seen with previous legislation, the environment can change which may open up loopholes in which boundaries can be pushed. Hence, solutions that offer a more thorough control that can be managed by the users themselves is therefore preferable. Furthermore, as seen in the tests with the previous legislation, many people are not aware of their personal data rights. The Internet was created for open communication for anyone (Leiner et al., 1997; Sweeney, 2015). Thus, it is difficult to make sure that every user comprehends all the legislations that covers the Internet.

This is an exploratory study of the proposed idea and further research and user studies would be required for increased significance of the findings. However, the solution addresses what was interpreted as the test participants' most pressing concerns about personal data on the Internet. Furthermore, except for minor alterations in the design, the prototype tests showed that it was easier to track and control what happened to the personal data. That is, the security and transparency of personal data usage were increased.

The blockchain technology solves many of the problems that inherently exist on the Internet to do with transparency and trust. There are many application areas and like many technologies surrounded by hype, people seek to exploit it. That is why Blockchain technology will probably need some time to be further developed and the technology to mature. It is a reasonable to assume that blockchain in the near future will change many of the current conventions around Internet use in different ways.

References

Acquisti, A. Taylor, C., & Wagman, L. (2016). The econonomics of privacy. Journal of Economic Literature, 54(2), 442-492.

Blackshaw, P., & Nazzaro, M. (2006). Consumer-generated media (CGM) 101: Word-of-mouth in the age of the web-fortified consumer (2nd ed.). Retrieved May 1, 2019, from http://www.nielsen-online.com/downloads/us/buzz/nbzm_wp_CGM101.pdf

Cerf, V. G. (2010). Trust and the internet. IEEE Internet Computing, 14(5), 95-96.

ComRes. (2014). ICO. Annual track 2014. Retrieved March 29, 2018, from https://ico.org.uk/media/about-the-ico/documents/1043485/annual-track-september-2014-individuals.pdf

Datainspektionen. (n.d.). Enkla grunder i dataskydd. Retrieved March 30, 2018, from https://www.datainspektionen.se/globalassets/dokument/enkla-grunder-i-dataskydd.pdf

The Economist. (2015, October 31). The great chain of being sure about things. The Economist, 21-24.

European Union. (2016, April 27). Regulation (EU) 2016/679 of the European Parliament and of the Council. Retrieved May 1, 2019, from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Fontana, A., & Frey, J. H. (2005). The Interview: From Neutral Stance to Political Involvement. In N. K. Denzin, & Y. S. Lincoln (Eds.), The SAGE Handbook of Qualitative Research (3rd ed.) (pp. 695-727). Thousand Oaks, CA: Sage Publishing.

Girardin, F., Blat, J., & Ratti, C. (2008). Digital footprinting: Uncovering tourists with user-generated content. IEEE Pervasive Computing, 7(4), 36-43.

Gupta, V. (2017). The promise of blockchain is a world without middlemen. Harvard Business Review. Retrieved March 30, 2018, from https://hbr.org/2017/03/the-promise-of-blockchain-is-a-world-without-middlemen

Iansiti, M., & Lakhani, K. R. (2017). The Truth About Blockchain. Harvard Business Review, 95(1), 118-127.

Internet Society (2018). Your digital footprint. Internet Society. Retrieved April 20, 2018, from https://www.internetsociety.org/tutorials/your-digital-footprint-matters

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G., & Wolff, S. (1997). Brief history of the internet. Internet Society.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Retrieved March 10, 2018, from https://bitcoin.org/bitcoin.pdf

Norman, A. T. (2017). Blockchain technology explained. Amazon Digital Services LLC.

OECD. (2007). Participative web: User-created content. Retrieved May 1, 2019, http://www.oecd.org/dataoecd/57/14/38393115.pdf

Rainie, L., & Anderson, J. (2017). The fate of online trust in the next decade. Pew Research Center. Retrieved March 10, 2018, from http://www.pewinternet.org/2017/08/10/the-fate-of-online-trust-in-the-next-decade

Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018, March 17). How trump consultants exploited the Facebook data of millions. The New York Times. Retrieved June 10, 2018, from https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Statista. (2018). Number of internet users worldwide from 2005 to 2017. Retrieved March 5, 2018, from https://www.statista.com/statistics/273018/number-of-internet-users-worldwide

Sweeney, S. (2015). Internet not designed for security, warns international expert. CIO. Retrieved from March 10, 2018, https://www.cio.com.au/article/569270/internet-designed-security-warns-international-expert

Warren, S., & Brandeis, L. (1890). The right to privacy, Harvard Law Review, 4(5), 193-220.

Zyskind, G. Nathan, O., & Pentland, A. (2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. 2015 IEEE Security and Privacy Workshops, San Jose, CA, 180-184.