Regaining Control Over Data and Privacy

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Stephen Shiao

Spring 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Sean M. Ferguson, Department of Engineering and Society

Regaining Control Over Data and Privacy

In a world of big data, companies collect more and more personal information on customers in order to run and grow their businesses in order to stay competitive. As more data is collected, companies are able to look very closely into peoples' lives – ads on Google may be items that you've only mentioned to others and Facebook may recommend people you've just met, even offline. Furthermore, companies have sold or shared data with other legal entities, namely other companies or the government, so users may not know who exactly has their personal information. However, according to a survey given out in the U.S. by Pew Research Center on privacy, 95% of users consider being able to control who has their personal information important (Madden & Lee, 2015). This means that some companies have kept sacrificing users' privacy in order to accomplish their goals, despite consumer opinion.

Using the social construction of technology as a framework, this paper explores the ways consumers can regain control and power over their own data and privacy, and how society may secure a right to privacy. I begin with an analysis on why privacy should be obtained through rules and regulations on companies. Then, I analyze two policies that have been enacted recently: Europe's General Data Protection Regulation and the California Consumer Privacy Act of 2018. These two were chosen because they are policies that detail specific handling of personal information, they affect all residents of Europe and California, and companies that do business with consumer personal information in Europe or California are affected, even if those companies are not in Europe or California themselves. By doing so, I could investigate how these acts can improve and new acts could be written to secure a right to privacy. This paper will not look closely into data vulnerabilities or security concerns and security concerns or ways to protect data from malicious hackers, those who access data with no permission.

Literature Review

Data is just information and can be classified in many different ways. This paper focuses on personal data, information specifically about an individual. There are four types of data associated with personal data (Sweeney, 2000):

1. Person-specific data: Details of information that are specific to an individual. This includes race, birthday, gender, or ZIP code.
2. Anonymous data: data cannot be manipulated or linked to confidently identify the entity that is the subject of the data
3. Explicit identifier: A set of data elements that provide a direct communication method with the person, so that they can be directly and uniquely contacted. This could include a name and an address or a name and a phone number.
4. De-identified data: data where all explicit identifiers are removed or replaced or generalized

The term big data refers to "information that could not be processed using traditional tools or processes" (Salvador & Ikeda, 2014). This came about from the sheer amount of information that is being exchanged on the Internet, with businesses like Twitter generating and collecting terabytes of data a day. There are three main steps businesses aim to achieve with big data:

1. Collecting and integrating data for fresh insights
2. Generate analytical models to automate operations and predict outcomes of business decisions
3. Create tools to take the results of their analytical models and take action, as well as train employees to use the tools

Following these steps, businesses have generated a digital trail for many people, complete with a profile and each individual's preferences, to attempt to predict a user's behavior. This helps them improve their product and marketing in order to attract more users, and ultimately generate more revenue.

Consumer Ability to Protect Privacy and Anonymity on the Internet

Consumers can protect their privacy and anonymity through several means (Porup, 2020). Some messaging services, like Signal, use encryption so that users' messages are only readable by the sender and receiver, and no unauthorized third party can read the messages. Using Tor will allow a user to avoid metadata, information about data, collection, but is unfortunately not fully robust – many services can detect Tor's use and block access. Other ways revolve around the user being careful about what information is given to services, like using zero-knowledge services so the service doesn't read what a user writes or checking app permissions so applications don't have access to every personal detail.

Unfortunately, these methods are not completely robust, and are too technical for most users. According to a study based on the 1990 U.S. census, just a few categories of person-specific data can uniquely identify an individual (Sweeney, 2000). 87% of the U.S. population was able to be uniquely identified with a 5-digit ZIP code, gender, and date of birth. 53% could be identified with a place, gender, and date of birth. 18% could be identified with a county, gender, and date of birth. Additionally, another study was done in the UK on network metadata (Alotibi, Clarke, Li, & Furnell, 2016). With 2 months of network data on 27 participants, the study was able to identify all of the participants. Even with de-identified data, it isn't difficult for users to be identified. Thus, if any of the aforementioned techniques in protecting privacy and anonymity fail, a user could be easily exposed on the Internet.

Some businesses try their hardest to get your data, no matter where you are. In the case of Facebook, they've implemented "like" and "share" buttons that other websites can add so that users can share what they've viewed. However, these buttons also collect metadata of the website's visitors (Lindsey, 2019). Since so many websites have these buttons, it isn't difficult for Facebook to formulate a profile for an individual, even if they don't have an account on Facebook. Other companies do this too – many news sites like CNBC or the New York Times have Facebook, Twitter, and LinkedIn share buttons. Considering the research on network metadata, it is impossible to know what entities have an individual's personal data and just how much.

Social Construction of Technology

In order to make privacy more accessible to consumers, companies themselves must provide that functionality. As such, there are now policies that regulate companies in the field of privacy and how they must handle personal information. To analyze the effect of these policies on different stakeholders, this paper makes use of the social construction of technology (SCOT) framework. SCOT focuses on the idea that the construction of technology is shaped by different social groups. These social groups look at technological artifacts different, and give them some positive or negative meaning that can influence the design of a technology in a certain way, either through design or stopping innovation completely. SCOT consists of four key components (Bjiker, Hughes, & Pinch, 1987):

1. Interpretative Flexibility: Every technological artifact has a different meaning for the different stakeholders
2. Relevant Social Groups: A group of people whose attached meanings to a specific artifact is the same. There are several of these for a given technological artifact, like users or

producers and third-party groups like policymakers. Each of these groups can be broken down into smaller subgroups, like different demographics of users.

3. Problems and Conflicts: Problems are the issues that some relevant social groups may face due to incompatibility. Conflicts are issues that may occur between some relevant social groups. SCOT analyzes these and connects them to the design features of a given technological artifact.
4. Closure and Stabilization: Closure is when a design fixes a problem or conflict. There are two types of closure:
    a. Rhetorical Closure: A design becomes good enough to satisfy all social groups.
    b. Redefinition of the Problem: Stabilizing conflicts by introducing a new problem that the design can solve.

    Closure is not permanent – a new social group could form and introduce new problems or conflicts. A closure stabilizes when it stays for a long term, and influences future innovation.

Relevant social groups for personal data include the owners of the personal information, in other words the consumer, the businesses, businesses that obtain the personal information from third-parties and not the user, and policymakers that are concerned with data and privacy. For the consumer, their interpretation of personal data is that it's their data – they should be able to control who knows or has the information. Consumers would also be able to withhold some information, possibly sensitive information that they wouldn't want other people to know offline. For businesses that directly deal with the customer, it is a valuable commodity that they can use to run the business and earn revenue and even sell for more revenue, aggregate to know their customer demographic, and analyze to better understand their customers, how they interact with

the business, and how they can improve their service. For businesses that get data from other entities, they use the data for the same things, but may not have a direct service to provide to the related customer. Policymakers are the third party that may or may not be related to the producer or consumer sides, but will have to step in in the event there is a problem or conflict that a policy would help solve.

The main conflict that exists is that businesses are collecting as much data as they can and sharing that data with other businesses, sacrificing consumer privacy. Some companies are even lobbying against giving more privacy rights to consumers because it would set back the business (Cadwalladr & Campbell, 2019). On the flip side, 95% of consumers value control over their personal data, but have low confidence that businesses will actually keep their data private and secure (Madden & Rainie, 2015). With little power over how companies continue to handle data, consumers must turn to policymakers to come up with a method of closure. Europe has come up with the General Data Protection Regulation and California has come up with the California Consumer Privacy Act.

General Data Protection Regulation

Implemented in 2018, the General Data Protection Regulation (GDPR) was pushed in the EU in 2016 to update their older legislation, one in 1950 that stated the right to privacy, and one in 1995 to establish bare minimum data privacy and security standards ("What is GDPR", N.D.). Google was then sued for scanning a user's emails, and Europe's data protection authority called for a more "comprehensive approach on personal data protection". Thus, the GDPR was created, affecting all entities that "process the personal data of EU citizens or residents, or offer goods or services to such people".

The GDPR outlines the following definitions ("What is GDPR", N.D.):

- Personal Data: Primarily the person-specific data, explicitly identifying data, and de-identified data as defined previously in this paper. Anonymous data can be included if it is "relatively easy to identify someone from it".
- Data Processing: Any action performed on data, whether automated or manual.
- Data Subject: The individual whose data is stored.
- Data Controller: The individual who decides why and how personal data will be processed.
- Data Processor: In the event the data controller doesn't actually do data processing, the third party that processes data for the data controller

There are seven general principles that data processors must follow under the GDPR. First, processing must be lawful, fair, and transparent to the data subject. Second, data must only be processed for legitimate purposes clearly stated to the data subject. Third, only necessary data for specified purposes should be collected. Fourth, all personal data should be accurate and kept up-to-date. Fifth, data can only be stored for as long as it takes for the specified purpose. Sixth, in all stages of processing, security, integrity, and confidentiality should be kept. This could be done through requiring employees to use two-factor authentication or end-to-end encryption. Finally, the data controller should be able to demonstrate GDPR compliance.

Even when in possession of personal data, there are specific times when any processing can be done on data. The data subject must have given clear consent to process the data in the first place. Then, processing is allowed when: it is needed in order to enter a contract where the data subject is a party, it is needed to comply with a legal obligation, it is needed to save

someone's life, it is needed to perform a task for public interest, or there is a "legitimate interest" to process a someone's personal data, like if the subject needed to be investigated. The last point is flexible, so the "fundamental rights and freedoms of the data subject" take precedence before the "legitimate interest".

Consent rules are also outlined, with the more relevant ones listed here. A data subject must clearly and willingly give consent, and any requests to a data subject must be unambiguous. Data subjects must also be able to withdraw any consent give at any time. Finally, evidence of consent must be documented. After data subjects give consent and data, the GDPR gives them many rights to privacy and data. Subjects now have the right to be informed of the processing that a company does, and if they don't like some parts, they can restrict or object to processing. They can also access all of the data they have given, and to be able to fix or delete any data. Subjects can also request the data so that they can move it around themselves – like giving it to another entity. They also have the right not to be subject to a decision made from automated processing or profiling, unless it's necessary for the company, authorized by the government with legitimate interests, or based on previously given explicit consent. Finally, if a data breach occurs, companies must notify affected users within 72 hours.

After GDPR, affected companies must work to be GDPR compliant. Primarily, they must add notices within their privacy policies clearly informing consumers of their rights given in the GDPR. They should also be clearer about getting explicit consent from consumers. Furthermore, companies must now designate a data protection officer, who specialized in GDPR compliance and data security strategy. A survey found that 68% of U.S.-based companies project $1-10 million in spending to be GDPR compliant, and 9% expect to spend more than $10 million

(Sykes, 2018). Those who don't face penalties of up to €20 million or 4% of their global annual revenue ("What is GDPR", N.D.).

GDPR lead to many consequences for both producers and consumers. Consumers will have gotten all of the rights outlined in GDPR and have more meaningful interactions with the services they seek out, but companies may decide to react differently than expected by the EU government. Forbes outlines 15 unexpected consequences of the GDPR (Forbes Technology Council, 2018). This paper will examine the more set-in-stone effects. The GDPR has led to a wave of new regulations around the world, leading to more regulation of similar rights. However, similar to the terms and agreement that consumers check without reading, consumers check the box that informs them of GDPR rights without actually reading. Small businesses also were not able to spend the money to be GDPR compliant and may face fines and penalties, although the GDPR saw its intended consequences for larger companies. Free services may also begin to dwindle, since their revenue came from data about the users. Basic photography also got affected – many services like Instagram have users share images, but if the image contains other people, then every individual in the photo has the right to remove the photo. Independent developers also do not have the manpower to be GDPR compliant, so European residents simply may not get a product, resulting in less product availability. Similarly, U.S.-based websites may deny service to European residents. Cybercrime also became more difficult – with protecting consumer data, they also protect malicious hacker data.

With these new problems that arose, it is clear that the GDPR as closure is not sufficient. Especially with it only being implemented in 2018, there is not enough evidence that it was successful. Overall, it seemed successful in giving users rights and forcing companies to care more about personal data. However, it seemed unsuccessful in that small companies and

independent developers are unable to keep up, and certain services are now unfeasible due to GDPR restrictions. Future innovations should not stabilize around this policy, but rather an adjusted policy in the future.

California Consumer Privacy Act of 2018

The California Consumer Privacy Act of 2018 (CCPA) was pushed, taking inspiration from the GDPR, after Alistair Mactaggart started a ballot initiative after being told by a Google engineer that "consumers have no idea just how much data online companies have collected on them" ("California Consumer Privacy Act", 2019). Putting in his own $3.5 million, the ballot initiative received more than 629,000 signatures, allowing it to be put as an issue on the November 2018 ballot in California. This act was pushed as a compromise between the ballot initiative and tech industry lobbyists, and will be effective January 1, 2020.

The CCPA applies to any businesses that collects personal information of California residents (Heimes, 2018). An entity is a business if one of the following applies: its annual gross revenue exceeds $25 million; it handles personal information of 50,000 or more consumers households, or devices; or it derives 50% or more of its annual revenue from selling consumers' personal information. Additionally, it must satisfy all of the following for the CCPA to apply: it is a legal entity that operates to profit; it collects consumer personal information or has a third party collect it for them; it itself determines the purposes and means of processing consumer personal information; and it does business with a California resident. It also defines personal information as: "Information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household" (Section 1798.140(o)(1) in the CCPA).

The CCPA gives consumers rights over their privacy and businesses more regulation over handling of personal data (Matheson, 2018). Consumers get the right to request their data, the right to opt out of the selling of their data, and the right to deletion of their data, and businesses must clearly inform consumers of their rights to do these within their privacy policies. Furthermore, they must include two or more designated means for consumers to submit requests, e.g. through a toll-free number or online platform, and cannot require consumers to create an account to do so. However, they must be able to verify that a consumer is who they claim to be, and this could be through an account login. Upon request, businesses must disclose the categories and details of personal information the business has collected about the subject within the past 12 months, where they got it and where they have given it within the past 12 months, and the purposes of collecting or selling the data. This should be done in a clear, "readily usable format" for the consumer. However, they are exempt if a consumer exercises these requests too frequently, in which case they are allowed to deny or charge the consumer. Additionally, businesses can deny deletion in certain circumstances that are near identical to the circumstances outlined in the GDPR (Kessler & Rudawski, 2018). Finally, businesses may not be discriminatory towards consumers who exercise their requesting or deletion rights.

Overall, how businesses prepare for the CCPA will be similar to preparing for the GDPR (Heimes; 2018, Matheson, 2018). Businesses must be able to identify the source of personal information, what the data is, where the data is stored, and where it has gone. If the data is sloppily stored then they may face fines from the CCPA, and will not be able to give out an accurate aggregation of personal information to consumers that have requested it. Businesses should also be more vigilant in tracking their purposes with data, and track all parties who

receive the data. They may also train employees or hire a specialist to help consumers exercise their rights and to be CCPA compliant.

Longer-term consequences are still unclear for the CCPA, but may have similar ones to the GDPR. Since the CCPA was effective January 1, 2020, there is not enough data to determine set-in-stone consequences. It will have similar effects as the GDPR, as outlined previously. However, the CCPA affects larger businesses, whereas the GDPR affects basically all businesses. This means that the aforementioned small businesses that couldn't keep up with the GDPR could keep up with the CCPA, since it doesn't affect them. The CCPA was a step in the right direction as a rhetorical closure, solving the small business issue with the GDPR, but many of the problems that arose with the GDPR still persist within the CCPA and need to be fixed.

Conclusion

Through this research, it's evident that users themselves will have a very difficult time maintaining privacy by themselves, and that policymakers need to step in in order to remedy this. However, the policies that are being put into effect are too strict, and inhibit innovation from small companies and the different types of services that companies may produce. The desired effect on the large companies that knew too much about individuals' lives was achieved, but adverse side effects were brought about. Thus, new forms of closure should be created, either through the improvement of current policies, or the creation of new policies, keeping the previous ones in mind, for states that do not have modern privacy policies.

Bibliography

"2019 Consumer Data Privacy Legislation." Accessed February 26, 2020. https://www.ncsl.org/research/telecommunications-and-information-technology/consumer-data-privacy.aspx.

Alotibi, Gaseb, Nathan Clarke, Fudong Li, and Steven Furnell. "Identifying Users by Network Traffic Metadata." *International Journal of Chaotic Computing* 4 (December 1, 2016). https://doi.org/10.20533/ijcc.2046.3359.2016.0013.

Altman, Micah, Alexandra Wood, David R. O'Brien, and Urs Gasser. "Practical Approaches to Big Data Privacy over Time." *International Data Privacy Law* 8, no. 1 (February 1, 2018): 29–51. https://doi.org/10.1093/idpl/ipx027.

Bijker, Wiebe E., Thomas P. Hughes, and Trevor J. Pinch, eds. The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. Cambridge, MA: MIT Press, 1987.

Cadwalladr, Carole, and Duncan Campbell. "Revealed: Facebook's Global Lobbying against Data Privacy Laws." *The Observer*, March 2, 2019, sec. Technology. https://www.theguardian.com/technology/2019/mar/02/facebook-global-lobbying-campaign-against-data-privacy-laws-investment.

"California Consumer Privacy Act: Everything You Need to Know About CCPA, the New California Data Privacy Law." Accessed February 26, 2020. https://www.fairwarning.com/insights/blog/california-consumer-privacy-act-of-2018-everything-you-need-to-know-about-the-new-california-data-protection-law.

Data Protection Report. “CCPA Extends ‘Right to Deletion’ to California Residents,” September 27, 2018. https://www.dataprotectionreport.com/2018/09/ccpa-extends-right-to-deletion-to-california-residents/.

Forbes Technology Council. “Council Post: 15 Unexpected Consequences Of GDPR.” Forbes. Accessed March 7, 2020. https://www.forbes.com/sites/forbestechcouncil/2018/08/15/15-unexpected-consequences-of-gdpr/.

“Four Strategies to Capture and Create Value from Big Data •.” Accessed February 11, 2020. https://iveybusinessjournal.com/publication/four-strategies-to-capture-and-create-value-from-big-data/.

Haselton, Todd. “Facebook Explains How It Can Collect Info about You Even If You Never Post on Facebook.” CNBC, April 16, 2018. https://www.cnbc.com/2018/04/16/facebook-collects-data-even-when-youre-not-on-facebook.html.

Lindsey, Nicole. “EU Court Ruling on Facebook ‘Like’ Button Could Have Huge Implications for Websites.” CPO Magazine, August 16, 2019. https://www.cpomagazine.com/data-privacy/eu-court-ruling-on-facebook-like-button-could-have-huge-implications-for-websites/.

Madden, Mary, and Lee Rainie. “Americans’ Attitudes About Privacy, Security and Surveillance,” May 20, 2015. https://www.pewresearch.org/internet/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/.

Montjoye, Yves-Alexandre de, Laura Radaelli, Vivek Kumar Singh, and Alex “Sandy” Pentland. “Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata.” *Science* 347, no. 6221 (January 30, 2015): 536–39. https://doi.org/10.1126/science.1256297.

Pinch, Trevor J. and Wiebe E. Bijker. "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." Social Studies of Science 14 (August 1984): 399-441.

Porup, J. M. "8 Steps to Being (Almost) Completely Anonymous Online." CSO Online, February 11, 2020. https://www.csoonline.com/article/2975193/9-steps-completely-anonymous-online.html.

Rainie, Lee. "How Americans Feel about Social Media and Privacy," March 27, 2018. https://www.pewresearch.org/fact-tank/2018/03/27/americans-complicated-feelings-about-social-media-in-an-era-of-privacy-concerns/.

Rainie, Lee, and Janna Anderson. "Implications of The Internet of Things Connectivity Binge," June 6, 2017. https://www.pewresearch.org/internet/2017/06/06/the-internet-of-things-connectivity-binge-what-are-the-implications/.

Salvador, Alexandre Borba, and Ana Akemi Ikeda. "Big Data Usage in the Marketing Information System." *Journal of Data Analysis and Information Processing* 02, no. 03 (2014): 77–85. https://doi.org/10.4236/jdaip.2014.23010.

Sweeney, Latanya. "Simple Demographics Often Identify People Uniquely." . . *Pittsburgh*, 2000, 34.

Sykes, Nathan. "What Is GDPR and How Will It Affect Your Business in 2018?" TechTalks, February 2, 2018. https://bdtechtalks.com/2018/02/02/what-is-gdpr-how-will-it-affect-your-business-in-2018/.

"Top 5 Operational Impacts of CCPA: Part 5 - Penalties and Enforcement Mechanisms." Accessed March 5, 2020. https://iapp.org/news/a/top-5-operational-impacts-of-cacpa-part-5-penalties-and-enforcement-mechanisms/.

"Top 5 Operational Impacts of the CCPA: Part 1 — Determining If You're a Business Collecting or Selling Consumers' Personal Information." Accessed March 5, 2020. https://iapp.org/news/a/top-

five-operational-impacts-of-cacpa-part-1-determining-if-youre-a-business-collecting-or-selling-consumers-personal-information/.

"Top 5 Operational Impacts of the CCPA: Part 2 - Transparency and Notice Obligations." Accessed March 5, 2020. https://iapp.org/news/a/top-5-operational-impacts-of-cacpa-part-2-transparency-and-notice-obligations/.

"Top 5 Operational Impacts of the CCPA: Part 3 - Responding to Consumers' Personal Information Access Requests." Accessed March 5, 2020. https://iapp.org/news/a/top-5-operational-impacts-of-cacpa-part-3-responding-to-consumers-personal-information-access-requests/.

"Top 5 Operational Impacts of the CCPA: Part 4 — Rights of Erasure, Objection to Sale, and Nondiscrimination." Accessed March 5, 2020. https://iapp.org/news/a/top-5-operational-impacts-of-cacpa-part-4-rights-of-erasure-objection-to-sale-and-non-discrimination/.

GDPR.eu. "What Is GDPR, the EU's New Data Protection Law?," November 7, 2018. https://gdpr.eu/what-is-gdpr/.