Password Systems and Being Compliant with Article 25 of GDPR

Clifton Paul Robinson
Dr. Woodrow Hartzog, Esq.
Northeastern University
Khoury College of Computer Sciences
Spring 2020

Introduction

It is a well-known fact in the cyber community that the average Internet user is the primary security risk. Adversaries know that the easiest way to exploit online systems is by targeting the users through social engineering attacks. These attacks are intended to trick the user into giving information, allowing adversaries to gain access to sensitive information or make a monetary gain. One specific area of social engineering attacks deals with passwords. A study in 2017 showed that in the United States the average email address is associated with 130 accounts (Uncovering Password Habits...). Most, if not all, people could not remember 130 unique and complex passwords, and this leads to password reuse. When a majority of passwords lack complexity, it leads to easy prey for adversaries. Being a major issue, password systems have worked to improve security for the users.

Websites that include user accounts have to update their password systems to secure users' accounts, and prevent sensitive data from being stolen. As there are countless ways to add security to password systems, some common examples are character and symbol requirements, multi-factor authentication, and known devices or locations. Systems usually include one, if not all, of these methods. This is because it offers a resolution to accounts with weak passwords. While these methods are great at ensuring more protections against password-based attacks, there starts to be some ambiguity as to whether some of these methods violate Article 25 of the General Data Protection Regulation (GDPR).

The General Data Protection Regulation is a regulation that is in EU law that is based on data protection and privacy in the European Union and the European Economic Area. It was created in April of 2016 and was placed into effect in May of 2018. This regulation gives the

consumer more power and control when it comes to their personal data (Tamò-Larrieux). This also makes companies become compliant with the regulation and if a company fails to achieve that they can be fined for it. Focusing particularly on Article 25, this article focuses on how companies implement their products, making sure that the data collected is stored and used correctly. Also, it makes sure that all the personal data used for the product to work is necessary and not being used, or overused, for the wrong reasons (Tamò-Larrieux).

Password systems have developed to become more dependent on specific personal data to ensure that the security and integrity of the accounts stay intact. This data is important when it comes to securing accounts, but it is now unclear that if the data used in these systems violates article 25 of GDPR, data protection by design and by default. The reason it is unclear is that with GDPR there is still ambiguity based on how new this regulation is phrased. This is especially true for Article 25 because companies will argue the importance of the data they are collecting for protective measures. The goal of this paper is to look at how Article 25 and current password systems should work together and offer clarity as to how we advance when looking for GDPR violations and protecting private data for citizens of the European Union (EU) while not risking the security of their online accounts.

Offering clarity to parts of GDPR is not always that simple, as the GDPR was adopted on 14 April 2016 and has only been enforceable since May 2018. This means there are only a handful of violations to look at and even less that are based around this subject in general. However, by looking at violations that have already been given, guidelines for being GDPR compliant, and other work based around this subject, a conclusion can be made that will offer more clarity and guidance for these companies to follow when working with these systems.

Confusion and Compliance within GDPR

It comes as no surprise that companies have found complying with GDPR to be difficult. Some companies have even stated they feel somewhat unprepared for GDPR, only being able to use temporary resources to fit the needs until they can find the best solution for themselves (Mikkelsen). This is especially true for companies that are no longer compliant in systems they have been using for years. Therefore, issues may arise that could potentially blindside a company based on what they currently have implemented.

The GDPR offers general guidelines to these organizations that offer them information about becoming compliant; however, the issue is that these explanations are too broad. To a certain extent, these companies are updating their systems based on partial information and they usually are not equipped to update everything at once to become compliant. Another dilemma that arises is the risk taken when updating everything at once. If organizations attempt to reach full compliance at once, they risk creating new vulnerabilities, overlooking existing vulnerabilities. This could also lead to systems failing to communicate because of the quick turnaround. The best recommendation for these companies is to implement solutions in a more controlled way through internal reviews and audits (Mikkelsen).

The GDPR has created a steep learning curve for everyone involved. This is why EU regulatory offices have had difficulties filling positions, creating new challenges when attempting to keep up with the demands of the regulation. Consumers have been given much more power in the information they have online and have had to adapt to privacy pop-up notifications online everywhere they go. Lastly, companies have had issues with the new internal data bureaucracies and adjusting to the vast changes required for compliance. With the learning

curves of this regulation, it is also important to look at the specific articles to gain an understanding in the broad sense.

An Introduction to Article 25 of the GDPR

Article 25 is now the way that user data is protected in the EU, including how it is obtained and how it is distributed. In other words,"Data protection by design and by default." The detailed explanation of what Article 25 of GDPR includes is,

> The controller shall implement appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage, and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons (GDPR, Article 25).

This means that companies can only use and collect the data that is absolutely necessary when it comes to their products and systems. It also means that users need to provide consent to a company to allow their personal data to be used and given out.

Article 25 is one of the groundbreaking parts of the GDPR because it gives personal data back to the user. This article offers users a strong first step in obtaining digital civil liberties, helping change the landscape of data collection overall. There now needs to be consent given to use that data outside of what they said they would do. It also allows users to pick and choose

which data a company is allowed to collect, unless the company is able to prove that it is necessary data that is needed.

Even though Article 25 seems like it is black and white, there continues to be confusion around the regulations. The confusion comes from the line, "only personal data which are necessary for each specific purpose of the processing are processed." Now the question of what constitutes necessary personal data and consequent answer to that question varies on a case-by-case basis. This is where we find a grey area that makes it difficult for companies to decide what data is necessary and what may violate Article 25.

The European Data Protection Board (EDPB) has published guidelines on how to stay compliant with Article 25 and they are helpful in understanding the scope of what needs to be done. It works more like an umbrella for everything rather than offer specific guidelines based on the type of system or product. No one should expect the EDPB to come out with guidelines for every type of system that relies on data. There are already numerous systems and new ones will continue to be invented. However, systems that are widely used and considered commonplace should have specific guidelines. This is due to the fact they will not be easily replaced in the near future.

Today, if you have an account online you will have an email and password attached to it. With the internet being a lawless place, having these passwords safe is a top priority. Hence, there is a strong GDPR password policy that helps guide companies to be compliant. These guidelines mainly focus on the password itself, not the entire system used for the login processes (GDPR Password Policy). This does offer users and providers a way to better protect themselves while forcing weak passwords out of the system. Therefore, the guideline itself falls short.

It is easy to classify emails, usernames, and passwords as necessary personal data. Without the three of those, no one can have an account that is protected and secure. The issue with this is that it focuses on only one aspect of these password systems without looking at where most of the personal data is stored. We have seen an advancement in technology to help make us more secure when it comes to these password systems. People have used personal questions to help prove to a system they are who they say they are. We are starting to see systems that store the locations based on where a user logs in. There are multi-authentication methods that require a phone number to allow access to accounts. The list can go on if we start to look at other methods and biometrics. When do we start to ask if this personal data is necessary information for these systems or an overreach? The answer is now.

Determining what Data is Necessary

Password systems have advanced tremendously since the birth of the Internet. From a single username and password to numerous preventative measures built on top of IT companies have worked to bolster these logins. However, how much personal data is needed, and is there a limit to it? When building a secure system in cybersecurity, it is known that a system can never be 100% secure. The only way to guarantee 100% security is to turn the system off entirely (Ghosemajumder). For companies that have their product online, this is not possible. So, the next best option is to make the system too expensive for an adversary to break into.

In cybersecurity, the terms cost and expense are one way to determine how long or how much it would take to attack a system. These methods can be used to determine if a system is even worth attacking. For example, if a system is using a public-key encryption method such as RSA (Rivest et. al.), an adversary can *technically* decrypt the encryption. The catch – it would

take hundreds if not thousands of years. For an attacker, the cost would be too high to attempt that. Once a certain point is reached, an adversary will probably not even attempt to break into the system (DynaSis). This plays into both sides of the argument, security should always be a top priority, but there needs to now be a classification between necessary and unnecessary data.

It is important to look at specific password systems to understand the personal data in question and what is being collected. The password login systems for Google and Facebook will be the cases we explore in this paper. Unfortunately, both of these company's systems are closed-source,[1] but anyone can get a decent amount of information by looking around their settings and seeing what information these companies have stored.

Google has established itself as a powerhouse on the Internet. From being the world's main search engine to Google Drive and Gmail, most people use Google in come capacity everyday. The statistics show this as well. In 2019, Google announced that they had surpassed 1.5 billion accounts worldwide (Elias). This demonstrates a trust built between Google and their customers, and that this trust is easily broken so the security in place can be strong. How much personal data is taken from this trust though? It is important to look at what Google collects to understand your protection.

While browsing Google settings, users are able to see every device they have ever used with that account. With this comes the type of device and the web browser being used with it. Depending on what kind of device this is, it can show a location as well. Just from this, Google knows your preferred type of phone, browser, the general location of where a user is. On top of that, they offer multi-factor authentication and email alerts. Google uses multi-factor

[1] Closed-source systems use code that is proprietary and kept secret to prevent its use by others.

authentication by either sending a message or call to a users' phone number, or by using their personal app. This adds another piece of sensitive data that Google stores about their users.

This is just the information found while browsing a Google account, as there is a possibility of more data being stored and encrypted on top of what a user can see and remove. Currently, this is the list of what Google stores on it's users just for the login process:

- Type of device
- Type of browser
- Phone Number
- Location data

Facebook works in a similar way to Google, but they are even more of a closed book. They boast about their internal password hashing algorithm to securely store passwords. This seemed to come up after a report in 2019 showed that Facebook was storing hundreds of millions of passwords in plaintext[2] since 2012 (KrebsonSecurity). If you attempt to login to Facebook and enter an old password, they will tell you that that is not the current password and you changed it from that password in a specific year. Similar to Google, they also store your login information so you do not need to continuously log back in after one use. Facebook offers two-factor authentication for all accounts, but does not require it for any of them. This adds an additional source of security to users, they just need to be the ones to enable it. This is the list of what Facebook stores the following information from their user login process:

- Old passwords
- Device Data

[2] Text that has not been encrypted or modified to be secure, *i.e. password = password*

- Phone information (based on the user agreeing to multi-factor authentication)

When comparing Google and Facebook, they both offer one important security concept, but do not require it fully for their users – multi-authentication. This is an authentication method that works as a "prove it" for the users. A user would give the system a password and then the system would challenge the user by asking for evidence of one or more devices that have previously been confirmed by the system. This authentication method has become a game changer. In late 2019, Microsoft came out with a report stating that multi-factor authentication blocks 99% of account hacks (Marks). This adds a new layer to the discussion with Article 25 and these password systems. If 100% security cannot be achieved and the cost for an adversary is high enough to deter them, that means there is only so much personal data needed in these systems.

When it comes to fraud prevention, more data might be necessary. A login system can be the barrier between data being safe or stolen. Companies have started to take greater steps to securing their systems to make sure none of their users accounts are hijacked. The main way that these companies bolster their password systems is by using specific personal data that is collected from their users, such as saved device data. By adding this information it improves their systems by pairing specific user information with their accounts to add a uniqueness that makes it more difficult to hack into without being noticed.

Today, user accounts hold much more data than ever before. This means that a compromised account or data breach would be an expensive problem to fix. A weak password

system could lead to the loss of credit card numbers or even social security numbers based on the account. This means that there may be no limit to the amount of necessary data when it comes to password systems. In theory, the more security measures means the safer the account will be. Does this mean that password systems are immune to Article 25? Possibly. As is most outcomes with GDPR, it is usually on a case by case basis.

Going back to the fact that we can not obtain 100% security in a system, this debate will boil down where a company prioritizes safety versus privacy. Every company has different priorities; some may say that privacy takes precedence, and others will say that it is security. There is no one correct answer to this problem. The only incorrect answers are if you sacrifice too much security where privacy is at stake or that there is too much security where there is no privacy between the user and the system.

New Guidelines from Article 25 Violations

While there has not been a Article 25 violation based around password systems, we can still look at other GDPR Article 25 violations to draw conclusions and guidelines based on current uses in these systems. We will look at the major Article 25 violations of the past two years to understand what the auditors are looking for.

In October 2019, a real estate company, Deutsche Wohnen SE was fined 14.5M euros by the Berlin Commissioner for Data Protection and Freedom of information. This fine was issued to them for lacking legitimate reasons to hold sensitive consumer data longer than necessary. The reason this fine was so steep was because they were cited for this issue in a prior complaint in 2017 and had the fine compounded (Berliner).

While this fine was extreme because of prior issues, it is still setting a precedent in fines for Article 25 violations. Stated earlier was how Facebook kept old passwords stored as well to remind users that that is not their password anymore. There does not seem to be any strong reasons to keep this information. If anything, it is more risky to have this information available. With the violation against Deutsche Wohnen SE in place, Facebook could be investigated for a similar violation based on how it stores any user passwords after it has been changed.

In December 2018, Facebook was fined 10M euros by the Italian Competition Authority for violating several articles of the Italian Competition Authority (ICA) as well as two articles of the GDPR, one of them being Article 25. The report found that Facebook actively sent consumer data to third-party sites and apps for commercial purposes. Facebook did this by default and without user consent. Facebook also made it harder for users to restrict consent, making their user experience harder. On top of that, they were also given a fine for their Cambridge Analytica scandal that is credited with helping influence the 2016 United States Presidential Election (Seals).

These actions clearly violate the core principle of Article 25. In this case, there was no data protection by design and by default – it was the exact opposite. This violation plays an important role in the foundation of Article 25; however, it would be hard to see any company abusing the information gathered from a password system to this level. If a company ever got to this level, there would probably be bigger issues than a GDPR violation here, potentially leading to election interference around the world, hurting current democracies.

The third violation was handed down by the Hellenic Data Protection Authority (HDPA) of Greece against a telecommunication service provider who ended up “punishing” the

customers that tried to cancel or transfer their service. The company sold their data to advertising companies even though consumers were told that their personal information was secure (edpb).

Similar to the previous Facebook violation, this is a type of violation you would hope you would not see when it comes to a password system, especially if the company was selling customers actual passwords. However, a company could still hold onto your information after a user cancels an account. If a user used multi-authentication, then the company would have access to a phone number and other resources. Between the last two violations, those would be the ones that would be most important to crack down, as it would encourage bad behavior towards users who are exercising their rights provided by the GDPR.

This is an area in the GDPR that needs to be expanded on. While there are guidelines for passwords, the guidelines need to be extended to work with the entire system, and not just the password itself. While there are not many Article 25 violations, each one does offer new guidelines for updating how to be compliant with the GDPR. In a perfect scenario, guidelines can be released so there never needs to be a Article 25 violation for password systems, but that is not, and never will be, the case.

Based off of these three violations, you can create some important guidelines for compliance and password systems:

1. **Remove data that is no longer relevant**

   Seeing a steep fine given to Deutsche Wohnen SE for keeping personal data for longer than needed with no strong reason to keep it could see companies in trouble for this as well. This means that when a user changes a password, username, email, or anything else that would play a key part in a password system, it needs to be removed quickly. There is

no reason to keep a user's outdated information stored unless it is being used to just keep as data on them. By removing any unnecessary or outdated data, a company's password system will stay compliant with Article 25.

2. **Keep specific contact information in a seperate database**

   Seeing two violations based on exploiting the personal data of their users is unnerving to think about. Everything is based on trust between the users and companies, and these are quick ways to break that. One way to achieve this is to have a specific database just for multi-factor authentication. This database would deal only with the security measure and not be touched for anything commercial.

It is difficult to look at the current Article 25 violations and create a long list of guidelines simply because there are not many to look at. However, guidelines can be added from the article itself in the future.

The problem itself is not a hard issue to solve, but there will never be one correct answer. The GDPR has a lot of grey areas, and the guidelines need to make sure that there is only a small amount that is left ambiguous. From the current Article 25 password guidelines, these would be strong guidelines that would help with compliance:

1. **Offer a base-level password system:**

   Just like how users can opt-out of what happens to their data, companies should allow them to opt-in to stronger password systems based on what personal data is used. This does not mean that companies use a skeleton password system as a base. A company's base password system will still be strongly secure, but it does not use any data that seems

unnecessary. This takes out a lot of the ambiguity of what is necessary data and also allows the user to add more security to their accounts by allowing the data they choose.

2. **Explain why the data used is necessary:**

   As stated multiple times, Article 25 is almost open for interpretation. If a company believes that the data they use is necessary, then they should state this to their users and the reasons for it being important in this system. By doing this, a company will be demonstrating their compliance willingness to comply. If a company ends up violating this article it would be easy to enforce a fine against them.

3. **Focus your login systems around multi-factor authentication:**

   The security results from multi-factor authentication are hard to ignore; however, ease of use will play a role in how these systems advance. Companies want to make sure that their systems are easy to use. Multi-factor authentication can be seen as tedious and annoying for users. However, by starting to normalize this in many password systems, users will get used to it. This has upside in multiple ways. First, it offers companies an easy and strong login system to implement and second, it helps fix the social engineering attack dilemma by adding an additional level of safety to the users as well.

These three guidelines would improve Article 25 compliance with password systems. While it would be easier to base more guidelines off of current violations, creating guidelines for these systems is going to need to stay broad because it will vary from every system.

Conclusion

While we wish that all the answers were here now, GDPR will continue to have a learning curve. Having a regulation that revolves around general data protection will never be

perfect, but it will need to stay resilient with the times. These password systems must continue to stay resilient with the times as well. In addition, companies need to keep finding new ways to stay up-to-date with these adversaries and provide strong protection for their users. In a sense, GDPR is just another step for improving online protection as well. Companies will need to either adjust their systems to address unneeded data or they will need to address exactly why it is necessary.

By addressing the problem of password systems, you are not saying anything is currently wrong with the system, but you address the ambiguity that is around the subject. If the European Data Protection Board expands their guidelines to focus on the entire system rather than just the users' password, then Article 25 of GDPR will have taken a step in the right direction to secure both the users' data and accounts.

Works Cited

(2019, November 5). Retrieved from https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/pressemitteilungen/2019/20191105-PM-Bussgeld_DW.pdf

Administrative fines imposed on a telephone service provider. (2019, October 7). Retrieved from https://edpb.europa.eu/news/national-news/2019/administrative-fines-imposed-telephone-service-provider_en

Another fine for violating the RGPD. (2019, December 10). Retrieved from https://www.dataprotection.ro/?page=Alta_amenda_pentru_incalcarea_RGPD_2020_1&lang=ro

Elias, J., & Petrova, M. (2019, October 27). Google's rocky path to email domination. Retrieved from https://www.cnbc.com/2019/10/26/gmail-dominates-consumer-email-with-1point5-billion-users.html

European Union, Parliament, European Data Protection Board (EDPB). "Guidelines on Article 25 Data Protection by Design and by Default." *Guidelines on Article 25 Data Protection by Design and by Default*, EDPB, 2019.

Fares. (2020, January 20). GDPR: The 6 Biggest Fines Enforced by Regulators So Far. Retrieved from https://secureprivacy.ai/gdpr-the-6-biggest-fines-enforced-by-regulators-so-far/

Fazzini, K. (2019, May 5). Europe's sweeping privacy rule was supposed to change the internet, but so far it's mostly created frustration for users, companies, and regulators. Retrieved from https://www.cnbc.com/2019/05/04/gdpr-has-frustrated-users-and-regulators.html

GDPR Password Policy: Critical Components. (2019, September 4). Retrieved from https://www.enzoic.com/gdpr-password-policy-critical-components/

Ghosemajumder, S. (2017, December 4). You Can't Secure 100% of Your Data 100% of the Time. Retrieved from https://hbr.org/2017/12/you-cant-secure-100-of-your-data-100-of-the-time

Green, A. (2020, March 30). Data Security and Privacy Lessons From Recent GDPR Fines. Retrieved from https://www.varonis.com/blog/security-and-privacy-lessons-from-recent-gdpr-fines/

How Much Does a Cybersecurity Attack Actually Cost? (2020, March 6). Retrieved from https://dynasis.com/2019/03/price-security-how-much-cybersecurity-attack-actually-cost/

Krebs on Security. (2019, March 24). Retrieved from https://krebsonsecurity.com/2019/03/facebook-stored-hundreds-of-millions-of-user-passwords-in-plain-text-for-years/

McGovern, William. *Privacy and Data Protection Law*. Foundation Press, 2016.

Mikkelsen, D., Soller, H., Strandell-Jansson, M., & Wahlers, M. (2019, July). GDPR compliance since May 2018: A continuing challenge. Retrieved from https://www.mckinsey.com/business-functions/risk/our-insights/gdpr-compliance-after-may-2018-a-continuing-challenge

Rivest, R. L., Shamir, A., & Adleman, L. (1978). A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. DOI: 10.21236/ada606588

S. K. Sood, A. K. Sarje and K. Singh, "Cryptanalysis of password authentication schemes: Current status and key issues," *2009 Proceeding of International Conference on Methods and Models in Computer Science (ICM2CS)*, Delhi, 2009, pp. 1-7.

Seals, T. (2018, December 11). Facebook Fined $11.3M for Privacy Violations. Retrieved from https://threatpost.com/facebook-fined-privacy/139824/

Tamò-Larrieux Aurelia. *Designing for Privacy and Its Legal Framework Data Protection by Design and Default for the Internet of Things*. Springer, 2018.

Uncovering Password Habits: Are Users' Password Security Habits Improving? (Infographic). (2018, December 14). Retrieved from https://digitalguardian.com/

X. Huang, Y. Xiang, A. Chonka, J. Zhou and R. H. Deng, "A Generic Framework for Three-Factor Authentication: Preserving Security and Privacy in Distributed Systems," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1390-1397, Aug. 2011.