Department: Head

Editor: Name, xxxx@email

Privacy Compliance: Technology can help it!

Abstract—From personal information controllers’ perspective, published privacy protection regulations bring some challenges to them. This article identifies six challenges to achieve privacy compliance and discusses technical solutions to them.

The introduction Personal information is playing a more and more important role in smart information systems (SIS for short), such as personalized recommendation, Wise Information Technology of 120(WIT120 for short). At the same time, the personal in- formation involved in SIS could also cause some serious ethical problems, including price discrimination and job discrimination. To pre- vent these problems, many countries have pub- lished privacy protection regulations [?], [?], [?]. In these regulations,personal information controllers are defined as the instituions or persons who control the personal information. After these regulations went into effect, some personal information controllers have been fined up to billions of dollars for failing to comply with privacy protection regulations. For example, Facebook was fined of 5 billion dollars in 2019 since it transferred personal information into Cambridge analytica with- out user consent[1](#_bookmark0), Google faced a fine of 50 million euros in 2019 for processing personal information for commercial purpose without

1h[ttps://www.theguardian.com/technology/2019/jul/24/](http://www.theguardian.com/technology/2019/jul/24/)

facebook-to-pay-5bn-fine-as-regulator-files-cambridge- analytica-complaint

user consent[2](#_bookmark1). Therefore, privacy compliance is becoming an important problem for per- sonal information controllers.

Asking help from law experts and privacy protection experts is the most direct solution to privacy compliance. However, manually supervising the business activities related to personal information is unpractical since it cost much human effort. In this situation, applying technologies to help personal infor- mation controllers automatically follow the privacy protection regulations is a more prac- tical choice. To apply technology to help per- sonal information controllers achieve privacy compliance, two questions should be answered: where to apply technology and what technol- ogy to apply?

In this article, we try to answer the above questions according to the General Data Protection Regulation (GDPR for short) [?], Personal Information Security Specification (PISS for short) [?] and California Consumer Privacy Act (CCPA for short) [?] published by European Union (EU), PRC and the USA

2h[ttps://www.reuters.com/article/us-google-priv](http://www.reuters.com/article/us-google-privacy-)acy-

france/france-fines-google-57-million-for-european- privacy-rule-breach-idUSKCN1PF208

IT Professional Published by the IEEE Computer Society © 2019 IEEE 1

respectively. These regulations are the most well-known privacy regulations in the main countries and areas of the world. For the first question, we define the personal information lifecycle, which consists of personal informa- tion collection, storage, transfer, processing and deletion. In these five stages, we iden- tify six technical challenges to comply with regulations through analyzing the personal information owners’ rights and the personal information controllers’ obligations. These six challenges are ’user consent’, ’user verifica- tion’, ’guarantee and quantification of the anonymity’, ’monitoring of personal informa- tion processing’, ’finding of the unnecessary personal information’ and ’Data Traceability’. For the second question, we will introduce some recent research work that tackles some of the challenges above. For the rest of the challenges, we will also discuss the potential solutions.

Overall, utilizing the personal information in their SIS could greatly improve the personal information controllers’ service quality and revenue. However, directly collecting and us- ing the personal information will cause serious legal punishment. Meanwhile, hiring experts to achieve privacy compliance is impractical. By identifying the six technical challenges and proposing potential technical solutions, we show the feasibility of using technologies to help personal information controllers utilize the personal information without violating privacy protection regulations.

# TECHNICAL CHALLENGES

While the privacy regulations in different countries may have different requirements in detail, they consistently regulate the collec- tion, storage, transfer, processing and deletion of personal information. We define the five stages as personal information lifecycle, which represents the phases personal information experience from collection to deletion. The stages of personal information lifecycle and corresponding challenges are shown in figure

?? with detailed requirements in different stages (red nodes represent lifecycle stages, green nodes represent concrete provisions and orange nodes represent technical challenges).

In this section, we will analyze the con- crete provisions, which regulate the actions of personal information controllers in lifecycle stages, to identify the six technical challenges we mentioned in section ??.

Personal Information Collection

When collecting personal information, the most necessary and important legal require- ment is acquiring user consent. Before describ- ing the situations of acquiring user consent, all three regulations required that the user con- sent notice should have a clear form and avoid technical jargons or legal jargons to make the notice understandable for users (GDPR Art.13, PISS Art.5.6, CCPA Art.2). Then

all the situations of acquiring user consent are detailedly regulated. The first one is the normal situation, which means when directly collecting personal information from adult users, personal information controller should ask the users for consent (GDPR Art.7, PISS Art.5.3, CCPA Art.2). The second situation is that when personal information is collected from the third party, the personal information controllers should ensure that users give their consent to access the personal information either by directly asking users or requiring the third party provide related materials(GPDR Art.14, PISS Art.5.3, CCPA Art.2). For the third situation, if personal information is col- lected from children, the personal information controller should ask the children’s guardian for consent, different regulations have different definition of a ’child’ (Europe is under 16, PRC is under 14 and the USA is under 13) (GDPR Art.12, PISS Art.5.5, CCPA Art.5).

As for the conditions of no consent required, we do not discuss them here as these condi- tions rarely happen.

After these regulations went to effect, some problems have emerged. From a macro per- spective, more than 35% of European websites still did not provide cookie consent notice to ask for users’ consent until May, 2018 [?]. Besides, even some websites have created the cookie consent notice, the form and content of notice might lead users to make agree de- cision [?], [?], which violates the transparency provisions. From a micro perspective, some big

collection

processing

storage

**Personal Data**

**Lifecycle**

deletion

transfer

Find Unnecessary Data

Traceability

Delete Unnecessary Personal Information

Deletion in third party

Right of delete

Recording of processing

User Veriﬁcation

Right of rectiﬁcation

Monitor

Right to restriction of processing

Right of access

Audit mechanism

No consent required

From children

Consisting purpose with declaration

From third party

Consent Notice

Sensitive information

security accessment

Normal

Anonymity

Commercial goal

Figure 1. Personal information lifecycle and corresponding regulations’ provisions, different colors represent different classes

companies have been punished seriously for collecting personal information illegally. For example, Youtube needs to pay a fine of 170 million dollars as it collected children’s per- sonal information without parental consent[3](#_bookmark2). Based on the above provisions and emerged problems, we identify the first technical chal- lenge to achieve privacy compliance: how to send an understandable and transparent user consent notice to an authoritative person (user himself or user’s guardian) before collecting users’ personal information?

Personal Information Storage

When personal information are stored, ex- cept for the traditional data security require- ments which have been identified and re- searched for many years, some users’ rights are regulated additionally. First, users have the right to access(GDPR Art.15, PISS Art.7.4, CCPA Art.3), which means users can request personal information controllers to provide

3https://zeenews.india.com/companies/google-

youtube-to-pay-170-million-penalty-for-illegally- collecting-data-on-kids-2232131.html

their personal information and related process information at any time. Once receiving the request, the personal information controller must respond to the request within a lim- ited time. Second, users have the right to rectify(GDPR Art.16, PISS Art.7.5), which means users can request to modify the wrong personal information about them stored by personal information controllers. Third, users have the right to delete(GDPR Art.17, PISS Art.7.6, CCPA Art.3), which means users can request the personal information controllers to delete the stored personal information be- longing to them. Any personal information controller who does not maintain these three user rights will face legal punishment. For example, Google was fined 8 million in 2018 dollars by Swenden government for violation of the right to erasure[4](#_bookmark3).

These rights are two-edged swords. While these rights help users preserve their privacy better, they also bring new security threats

4h[ttps://www.phonearena.com/news/google-fined-8-](http://www.phonearena.com/news/google-fined-8-)

million-in-sweden\_id122905

to personal information stored by controllers. These threats are dangerous and hard to resolve. On the one hand, when receiving the users’ requests, some personal information controllers might respond to the users’ re- quests without user identify verification mech- anism [?], [?]. Exploiting this vulnerability, malicious attackers could pretend as regular personal information owners to access, modify and even delete the others’ personal informa- tion easily. On the other hand, when verifying the users’ identity, to protect users’ privacy, personal information controllers should not require users provide extra personal informa- tion [?]. Therefore, the rights above bring a new technical challenge for personal informa- tion controllers in the personal information storage stage: how to verify users’ identity without additional personal information when these users send requests to the personal information controllers?

Personal Information Transfer

To transfer personal information to the third parties , personal information controllers should have permissions of users and gov- ernment(GDPR Art.44, PISS Art.8.2, CCPA Art.2). The permissions are determined by many factors(e.g.,sensitivity of personal infor- mation, the privacy protection level of the data receiver, political reason, etc.). Some of these factors can not be measured by technical methods. Besides, data transfer does not happen as frequently as the activities in other stages and manually evaluating the compliance degree of data transfer activities is practical. Therefore, the privacy compliance in this stage mainly depends on the law and technology experts’ help.

Personal Information Processing

Personal information processing is the most important stage of personal information life- cycle as insights and value are produced by processing personal information. In the big data era, data analytic technologies are so powerful that personal information controllers can portray a person comprehensively with his or her personal information and related data (e.g., purchase history, social network account

information). If personal information can be processed arbitrarily by personal information controllers, the users could face many privacy risks, such as job discrimination and price discrimination. These risks will violate users’ legal rights seriously. Therefore, the personal information processing should be regulated strictly, furthermore, achieving privacy com- pliance in this stage is very important but difficult for personal information controllers.

If personal information is processed for commercial goals, the result of the processing should not point to one specific person(GDPR Art.21, PISS Art.7.3). In the other words, others can not infer anyone’s personal infor- mation or attribute from the data processing results. Considering that the personal informa- tion controllers punished by governments cur- rently are mainly commercial enterprises, the provision is very critical. Besides eliminating the personal orientation of results, personal information controllers should also assess the impact of each data processing technology on personal information security (GDPR Art.35, PISS Art.10.2). The provision requires the personal information controllers to evaluate the anonymity of each data processing tech- nology.

As more and more attacks on popular data mining algorithms are being proposed and machine learning security is becoming an important topic in the security field. The above two provisions bring a technical chal- lenge to personal information controllers: how to guarantee and assess the anonymity when processing the personal information?

Besides keeping and quantifying the anonymity of the personal information pro- cessing, some provisions are also made to prevent personal information controllers from processing personal information in the wrong way. First, personal information controllers should process personal information for the consistent purposes with those declared in the user consent notice(GDPR Art.6, PISS Art.7.3). The provision prevents the personal information controllers from processing per- sonal information for purpose denied by users. In addition, users also have right to restrict the data processing(GDPR Art.18, PISS.7.3),

such as restricting the accuracy of the data processing. To guarantee the compliance with the above two provisions, personal information controllers are also required to record the process activities and establish audit mech- anisms (GDPR Art.28,30, PISS Art.10.5).

These records and audit mechanisms will be the basis of data protection authorities to make judgement.

As personal information might be pro- cessed by many employees in different depart- ments with a high frequency, the personal in- formation might be processed for the purposes which are not permitted by users, especially in big companies. Therefore, the above provisions bring a technical challenge to the personal information controllers: how to ensure that personal information is processed for the al- lowed purposes under users’ restriction?

Personal Information deletion

Personal information deletion is triggered in two situations: users request to delete or the personal information is unnecessary for the purposes declared in the user consent notice. The former one depends on users. For the latter one, the personal information con- trollers are required to immediately delete the personal information which is not necessary for the declared goals (GDPR Art.5, PISS Art.6.1). The personal information controllers who store personal information for too long will face serious legal punishment. In 2019, the Germany government fined a German real estate company Deutsche Wohnen SE 14.5 million euros as it did not delete the per- sonal information that is no longer needed for the user permitted purposes[5](#_bookmark4). The provision makes the personal information controllers face a new techinial challenge: how to find the unnecessary personal information?

When personal information deletion is trig- gered, personal information controllers should ensure that all the collected and stored per- sonal information is deleted, including these has been transferred to the third parties (GDPR Art.17, PISS Art.7.6, CCPA Art.3).

To delete the personal information thoroughly, tracing the personal information which has been transferred is critical for personal infor- mation controllers.

# CURRENT AND POTENTIAL SOLUTIONS

Although all the above privacy protection regulations are only issued for less than two years, privacy compliance is becoming a hot topic in the security field. Recently, some technical solutions from both academia and in- dustrial field have been proposed to help per- sonal information controllers to achieve pri- vacy compliance. Some of the above challenges thus have correponding technical solutions. Moreover, due to short history of privacy compliance problem, there still exist some challenges could be resolved by technologies but not yet worked out. In this section, we will introduce some existing solutions and discuss some potential solutions to each challenge identified in Section ??

User Consent

The first challenge consists of two sub- problems: making an understandable and transparent user consent notice and contacting the authoritative person.

For the first sub-problem, a machine learn- ing based method has been proposed to make long and complex privacy policies become short and condensed notes [?], so that the consent notice can be more understandable for users. Furthermore, Utz et al. and Matte et al. [?], [?] quantify the impact of consent notice form and content on the users’ cookie consent decision through a large-scale user study, the user study result will also help personal information controllers make a user consent notice more transparent.

For the second problem, Locality-Sensitive Hashing(LSH) [?] and blockchain technol- ogy [?] are potential solutions. LSH is a class of hashing methods which map simi- lar elements to the same value with a high probability and blockchain is a decentralized and distributed data record system. To receive

5h[ttps://www.h](http://www.huntonprivacyblog.com/2019/11/06/berlin-)un[tonprivacyblog.com/2019/11/06/berlin-](http://www.huntonprivacyblog.com/2019/11/06/berlin-) the user consent notice, personal information

commissioner-issues-fine-to-deutsche-wohnen-se/

owners (users themselves or users’ guardians)

should write their personal information ab- stract produced by LSH-like technologies on a blockchain system. When personal informa- tion controllers collect personal information, they can search the contact information of personal information owners on the blockchain system by hashing the collected data, and then send a consent notice to the corresponding users. However, the accuracy of LSH is not satisfactory and there are still some security problems in the blockchain system, such as de-anonymity. Therefore, more efforts are still needed to tackle the problem.

In addition to the existing acdemic work, a startup company securiti.ai[6](#_bookmark5) also applies the machine learning algorithms to help personal information controllers to find the personal information owners and request user consent automatically.

User Verification

In the second challenge, malicious attack- ers can steal and destroy the personal infor- mation stored by personal information con- trollers easily. This is because personal infor- mation controllers are lack of user verification mechanism [?], [?], [?]. Therefore, the second challenge can be viewed as an authentication problem. Authentication is a classical problem in the security field and various authentication mechanisms have been proposed so far, such as password-based mechanism and biometric- based mechanism. These authentication mech- anisms are proved to be powerful tools against the fake identity attacks. But directly applying these mechanisms is not enough as personal information controllers can not collect users’ additional personal information.

In this situation, Coline Boniface et al. [?] propose a scheme which helps users generate pseudo identifiers to verify their identity, so users do not need to provide their true identi- fiers in their right requests. Through detailed security analysis, they show that the scheme can verify users’ identity and do not need to collect users’ other personal information. Then the personal information controllers could pre- serve the users’ legal rights without leaking

6https://securiti.ai/

personal information and collecting additional personal information. However, the scheme still has a limitation: it depends on a trusted execution environment, such as users’ web browsers. How to release the dependence is an open question.

Anonymity

To tackle the third challenge, personal information controllers should guarantee and quantification the anonymity when processing the personal information. In the history of anonymity research, many classical privacy- preserving data processing mechanisms have been proposed to face the problem, from sim- ply protecting privacy when data publishing, to protect privacy when conducting complex data analysis [?], [?], [?], [?]. Among these privacy protection mechanisms, differential privacy [?] receives the most attention for its strong privacy guarantee and low time consumption.

Differential privacy protects the individ- ual’s privacy by limiting the impact of a single sample on the data analysis result, which means the presence or absence of a single person’s personal information will not influence the analysis result a lot. Therefore, the data analysis result will represent the group’s features rather than point to a specific individual. An individual’s privacy can thus be guaranteed in this way. Meanwhile, re- searchers also recommend differential privacy as a powerful tool for the machine learning systems to resist dataset inference attacks [?]. Besides keeping anonymity, differential pri- vacy mechanisms provide a parameter *ϵ* to quantitatively make trade-off between the ac- curacy of analysis and the degree of privacy. The parameter *ϵ* is proportional to the analysis accuracy and inversely proportional to the degree of privacy. In addition, as differential privacy mechanism only adds some noise to the processing intermediate results, it will not bring too much time consumption to the data processing.

As the theory of differential privacy is becoming more and more complete, many researchers are working on applying differ- ential privacy in real data analysis scenar-

ios. Recently, several differential private pro- gramming frameworks have been proposed to support the privacy-preserving data analy- sis [?], [?], [?]. With these frameworks, personal information controllers can process personal information with quantitative anonymity and do not need to understand the details of differential privacy.

Monitoring

The fourth challenge also consists of two sub-problems: restricting the purpose of the personal information processing and ensuring the data processing behaviors are restricted. Before the above privacy protection regula- tions are published, many approaches have been presented to restrict the purpose of the data processing and monitor the system behavior, such as system log analysis, purpose- based access control and information flow analysis [?], [?], [?].

After GDPR went into effect, some re- searchers try to adopt the existing approaches in personal information processing scenarios to comply with regulations. For the first sub- problem, David Basin et al. [?] propose a formal model which determines the data pro- cess purpose by analyzing dataflow of system interprocess communication, thus the personal information process purposes could be re- stricted. In Basin’s approach, the dataflow analysis accuracy is key to the purpose restric- tion efficiency. For the second sub-problem, first-order temporal logic (FOTL for short) is a system log monitor tool that formulates the data process requirement and automati- cally monitors the system behavior [?]. The data process requirements from regulations and users could be formulated by FOTL and then automatically monitoring the system behavior [?]. But the limitation of FOTL is that the transfer process between plaintext form requirement and FOTL language should be complete manually, which greatly limits the efficiency of monitoring. An automatic language transfer will greatly improve the monitoring efficiency.

Besides monitoring the system in real time, another solution is guaranteeing the compli- ance when design the data process model [?].

The main idea is automatically transforming the business-oriented data flow model to the privacy-aware data flow model to follow the regulations’ requirements such as purpose re- striction, or accountability. Thus, the personal information controllers could focus on the bussiness processing design without concern- ing about privacy issues. Overall, automat- ically monitoring and auditing data process behavior is receiving more and more attention and more and more technical solutions will be proposed.

Find Unnecessary Data

Finding the unnecessary personal informa- tion in the system means finding the personal information which is never needed for the data process purposes permitted by users. From a reverse aspect, to find the unnec- essary personal information, an approach is finding the necessary personal information in the system and the rest of personal infor- mation is unnecessary. Given some purposes, the necessary personal information could be found by auditing and analyzing the data flow in the system [?], then the rest of personal information stored by personal information controllers is unnecessary. After finding the unnecessary personal information, the per- sonal information controllers can take proper measurements, such as ask additional consent from users or delete the unnecessary personal information, to eliminate the legal risks.

Traceability

Digital watermarking technology is a po- tential solution to trace the data. Digital wa- termarking technology hides some encrypted information into data without reducing the utility of the data. The data with watermark can thus be traceable. Previous digital wa- termarking technologies include digital water- marking for images and text documents [?], [?]. These technologies were mainly used for copyright protection and media data broad- cast monitoring before. Actually, personal information could also be viewed as digital assets. When personal information controllers transfer personal information to the third parties, they can add watermark into the

transferred personal information to trace it. If personal information controllers need to delete the personal information, they can delete the transferred personal information according to the data track.

# CONCLUSION

In this article, we identify six technical challenges to achieve privacy compliance from personal information controllers’ perspective. We then discuss some current and potential solutions to these challenges to show that technologies can help to achieve privacy com- pliance.

In conclusion, privacy compliance is an important and big problem, what we have discussed is only the tip of the iceberg. There are still many problems to be solved by re- searchers.

# ACKNOWLEDGMENT

We thank A, B, and C. This work was supported in part by a grant from XYZ. 1993. (Thesis or dissertation)