Verifiable and Practical Compliance for Data Privacy Laws

Manu Awasthi
Ashoka University
manu.awasthi@ashoka.edu.in

***Abstract*—A number of governments have legislated privacy laws in recent years. The most prominent international one covering multiple nations is the General Data Protection Regulation (GDPR) of the European Union. Many national and local governments are in the process of tabling similar legislation.**

To be compliant with privacy laws, software companies providing Software as a Service (SaaS) have changed internal practices to develop applications with a "privacy first" ethos. In addition, these companies (data controllers) have put mechanisms in place for ensuring the privacy and security preparedness of their service providers (data processors), which is currently being done manually using questionnaires. Questionnaires designed to collect compliance information from processors aren't the best instruments. This is due to many reasons including lack of clarity on information to be collected, humans in the information collection loop, and badly designed questionnaires, among others. In this paper, we analyse a few reasons making compliance determination a herculean tasks for both parties and propose a simple mechanism to automate compliance information gathering.

I. Introduction

A number of privacy and data protection laws have been enacted by multiple governments at the national and local levels to protect citizen and customer data including the General Data Protection Regulation (GDPR) [1] and the California Privacy Rights Act (CCPA) [2]. GDPR in particular, provides a number of protections for the consumers of software services regarding their data, especially personally identifiable information (PII) collected on them. These include purpose limitation, providing data subjects rights to get data amended or deleted upon request, among many others. The data subjects have to be informed about types of data being collected about them, and consent has to be sought before any data is collected. GDPR also directs organizations to store and transmit user data with privacy guarantees. Many countries have enacted GDPR-like legislation, with some like South Korea providing more stringent protections [3]. Many others like India are in the process of enacting data protection legislation for their citizens [4].

Software organizations, especially ones providing Software as a Service (SaaS), are most affected by data laws since many of them *have* to store user data, to provide services and maintain state which helps provide a notion of continuity across sessions and devices. Many SaaS companies rely on third party SaaS providers for a subset of services they provide to their users. For example, an edtech company may embed content (videos, quizzes) which are designed, curated and hosted by a third party provider. Similarly, a news website may use third party services to provide its users access to online games and puzzles. In GDPR parlance, the main software service provider (like the edtech company or the news website) which collects data directly from citizens under a given law's jurisdiction is known as **data controller**, while the third party service provider is known as a **data processor** [5]. The data controller determines the purposes for which any consumer data is being collected, and also the means for doing so. As a part of compliance requirements, the duties of the processor towards the controller need to be laid out in legal contracts, making it the responsibility of the controller to make sure that the processor(s) are complaint with appropriate laws.

One data controller may utilize the services of many service provider or data processors. For example, the edtech company might be using one provider for hosting it's video content, but a different one for assessment services. Similar situations exist for multiple other SaaS providers (controllers) including apps to manage entry and exit for residents in apartment complexes and information management systems being used in universities – almost all require integration with multiple third party providers (processors). As the range of services being contracted out to processors increase, the burden of checking and verifying compliance of all processors across multiple jurisdictions on the controller also increases.

II. The Problem

Ideally, the controller would want to assess a processor's compliance by carrying out an extensive analysis of the deployment and development setup, including the data model and schema details of the processor's application. However, it's impossible to get such access – the processors would not want to provide it to protect their code, trade secrets and deployment details from being leaked.

In the absence of such access, the controllers have to settle for non-invasive mechanisms. There are typically two ways of doing it. The first is to ask the processor to provide audit and compliance check reports of their deployment done by a third party auditor. At times, certifications for ISO 27001 [6] or NIST CSF [7] serve as proxies for compliance to data privacy laws. The second and more common mechanism is to designate teams at controller for checking processors' compliance before availing their services. These teams typically comprise lawyers, HR professional, and members of the controller's IT and INFOSEC teams.

The compliance check is typically done by gathering information from processors through questionnaires. These questionnaires are designed to gather relevant information on information vectors deemed compliance critical by the controller. The responses to these questionnaires help the controller determine the efficacy of data storage, privacy, security and service availability practices of data processors. In essence, the controller looks for answers to many questions, some of which are presented below:

1) What data is being collected on users and does it include any PII?

2) Is user data being stored in a format that makes is amenable to be deleted completely, and off all copies of the data? How is this being ensured?
3) Is data encrypted at rest? Are there data storage security policies in place?
4) Is data secured during transmissions between application's microservices and to the user?
5) What are the data access policies of the organization? Who within the organization has access to user data? Is data access need-to-know?
6) What are the types of devices using which processors' employees can access user data? Is data on all such devices secure?

The controllers have to check for user information leakage from multiple sources at the processor. Not only they have to make sure that user information is secured at the primary source, i.e., the deployment(s), but also have to check other places from where user information could potentially *leak*. These include other devices (and hence people) within the organization. These considerations typically increase the scope of information being sought in these questionnaires – the controller has to seek information about the general security practices across the *entire organization* of the processor. As a result, the questionnaires become exhaustive to also include personnel policies and cultural practices within the processor's organization.

III. The Solution

Multiple issues arise in the process of collecting compliance related information via questionnaires. First, there doesn't exist an industry standard which can be followed to determine *all* the information that is to be collected by a controller – this remains a subjective decision of the controller's compliance determination team. Second, within a controller's organization, various teams might require different information from different processors, depending on the type of services provided by a processor. Additionally, depending on compliance jurisdiction, the information to be gathered from a processor may vary. Finally, the fact that one controller might rely on services of multiple processors and has to check for their compliance in every jurisdiction. All these result in creation of a single, very exhaustive questionnaire by a controller for *every* processor. And, the presence of multiple humans in the loop for designing compliance information collection questionnaires (from controller) and gathering and transmitting this information to the controller (from processor) makes the process tedious and error prone.

A part of this problem can be solved by the processor *voluntarily* providing a subset of compliance information for a given jurisdiction, and making it available publicly to any controller who wants to check their compliance for a given privacy law. This has to be done without compromising the processor's app deployment security or revealing any application details or trade secrets. One mechanism would be for the processor to create public API endpoints which can be queried by potential controllers for relevant information.

Each processor can maintain one endpoint that can be queried to return an object with basic compliance information for a jurisdiction. The returned object can contain answers to most common concerns around at-rest data encryption, data encryption while being transferred between deployment servers and information regarding collection of any PII. It could also mention the versions of s/w libraries being used by the processor so that the controller can check if any vulnerable library versions are being used. Other information might include active certifications, presence of a data privacy officer etc. Providing the first tranche of compliance information in this fashion reduces humans in the loop, since data is being provided automatically. It also helps cut down the time required for making a first cut assessment of a processor's compliance status.

Of course, it is not expected that *all* compliance information could be supplied in this fashion. If there are more details required which cannot be provided in an automated, public fashion, they can be supplied later by information collection via questionnaires. However, the benefit of having a two pass process like this makes sure that the questionnaires being designed for the second pass for a processor are tailored to the information being sought from *that* processor. Since this will be a targeted questionnaire based pass, with a lot less information to be collated, the small number of questions can be specific to the processor, and can embed the context for the services, deployment and application development models etc of every processor.

IV. Summary and Conclusions

In this paper, we highlighted the issues that arise between data controllers and data processors while exchanging information for verifying the compliance of processors to data privacy laws. We posit that the current process which involves multiple humans in the loop for creating mechanisms for collating compliance information, as well as parsing collected information to guarantee processor's compliance are error prone. To counter this, we propose a simple two pass process of exchanging jurisdiction specific compliance related information between the controller and the processor. The first pass allows for an automated querying of compliance information that a processor provides (via an Internet endpoint possibly) by the controller, after due authentication. This pass provides the controller with basic compliance information. If this information satisfies the compliance requirements of a particular jurisdiction, the entire process of human in the loop can be avoided. Even if this doesn't provide complete information, in the subsequent passes the controller can provide a small number of targeted, context-specific questions to individual processors to gather all required information, reducing the chances of error or incomprehension.

References

[1] "General Data Protection Regulation (GDPR) Compliance Guidelines," https://gdpr.eu/, Accessed: 2022-10-13.
[2] "California Consumer Privacy Act (CCPA)," https://oag.ca.gov/privacy/ccpa, Accessed: 2022-10-13.
[3] "16 Countries with GDPR-like Data Privacy Laws," https://securityscorecard.com/blog/countries-with-gdpr-like-data-privacy-laws, Accessed: 2022-10-13.
[4] "Govt looks to table data bill soon, draft at advanced stage," https://economictimes.indiatimes.com/tech/technology/govt-looks-to-table-data-bill-soon-draft-at-advanced-stage/articleshow/93355126.cms, Accessed: 2022-10-13.
[5] "What is a data controller or a data processor?" https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/controller-processor/what-data-controller-or-data-processor_en, Accessed: 2022-10-13.
[6] "ISO/IEC 27001 Information security management," https://www.iso.org/isoiec-27001-information-security.html, Accessed: 2022-10-13.
[7] "Framework for Improving Critical Infrastructure Cybersecurity," https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdfl, Accessed: 2022-10-13.