We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Modern Privacy Threats and Privacy Preservation Techniques in Data Analytics

Ram Mohan Rao P, S. Murali Krishna and AP Siva Kumar

Abstract

Today we are living in a digital rich and technology driven world where extremely large amounts of data get generated every hour in the public domain, which also includes personal data. Applications like social media, e-commerce, smartphone apps, etc. collect a lot of personal data which can harm individual privacy if leaked, and hence ethical code of conduct is required to ensure data privacy. Some of the privacy threats include Digital profiling, cyberstalking, recommendation systems, etc. leading to the disclosure of sensitive data and sharing of data without the consent of the data owner. Data Privacy has gained significant importance in the recent times and it is evident from the privacy legislation passed in more than 100 countries. Firms dealing with data-sensitive applications need to abide by the privacy legislation of respective territorial regions. To overcome these privacy challenges by incorporating privacy regulations, we have designed guidelines for application development, incorporating key features of privacy regulations along with the implementation strategies which will help in developing data-sensitive applications which can offer strong and coherent privacy protection of personal data.

Keywords: Data privacy, ethical code of conduct, privacy legislations, privacy preservation, design strategies

1. Introduction

Broad use of smart phones, e-commerce, social media, Internet and Communication Technologies (ICT) has transformed our lives. Though digitization facilitates our work, it is prone to privacy vulnerabilities. The key privacy threats include surveillance, disclosure, targeted advertisements [1], identity theft, information disclosure without consent, personal abuse through cyber stalking [2], studying emotions and mood of the people by accessing profile pictures, tweets, likes and comments to find emotionally weak, people who are lonely and trap them using various cyber-attacks like ransom ware, sexual abuse etc. [3]. Firms dealing with data sensitive applications need to follow certain ethical code of conduct to ensure privacy and guard the users from various digital assaults.

Digital data include variety of personal data like transactional data, location data, electronic medical records, e-commerce data, insurance data, photos and videos, opinions and views etc. All these data items are personal and sensitive data and should not be disclosed without the consent of the data owner. Privacy breach can occur at various stages of data processing as described in **Figure 1**.
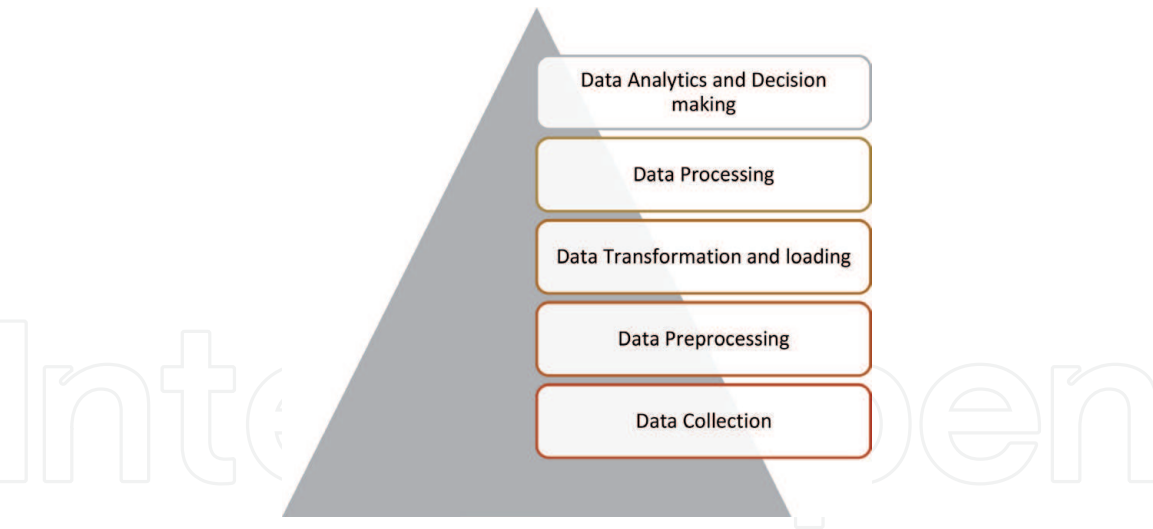
Figure 1.
Data processing stages.

Data breach can occur at any stage of data processing with different type of people operating at various levels. The top-level management should ensure that no data breach occur at any stage for which there is a need to have a policy in place and ethical code of conduct for all employees of the firm. However, policies alone are not sufficient; there must be regulatory body to ascertain the policies are implemented. Apart from this, the individuals are also contributing to data leakage by inappropriate use of social media and smart phones. Hence there are three entities responsible to ensure privacy preservation (**Figure 2**).

Governments and regulatory bodies are more responsible than others because Governments can impose privacy regulations and make sure the data holder or data collecting firms abide by them. Data holders are also equally responsible because data is with them and they can share the data with third parties without the knowledge of the data owner. By inappropriate use of social media applications like Facebook, Instagram, etc. users are also uploading personal data into the public domain which leads to privacy threats. With the consistent escalation of privacy threats and their grave consequences, awareness among users has also increased and in turn increased the demand for privacy preservation, which eventually led to the creation of privacy laws and regulations in many countries. The most prominent among them are GDPR (General Data Protection Regulation) of the European Union and the Personal Data Protection bill of India. Some of the applications along with its privacy risk are listed in **Table 1**.
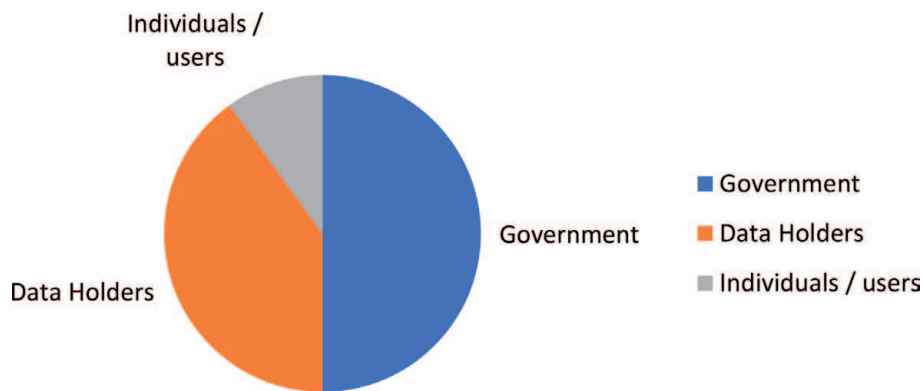

Figure 2.
Entities responsible for privacy preservation.

S.no.	Application type	Privacy risk involved
1	Smart phone apps.	Information theft, Intrusion
2	e-Commerce sites	Inference attacks, Disclosure
3	Social media	Cyber stalking, Ransom ware
4	Data capturing systems like banking, hospitals, insurance, government portals etc.	Disclosure, Discrimination

Table 1.
Application vs. privacy risk.

2. Ethical code of conduct through privacy legislations

Incredible amount of digital data is generated by virtue of various applications and technologies. This digital data will also contain personal and sensitive data of an individual which must be confidential, secure and private. Data privacy is the responsibility of the firms that capture the data and ensure no privacy breach in any stage of data processing. Hence there is a need for ethical code of conduct for privacy preservation of personal and sensitive data in private and public domains. **Figure 3** depicts few important practices of ethical code of conduct in data processing.

There is a need for privacy legislations because of modern privacy threats and also to ensure ethical practices being followed by data holders. Some of the modern privacy threats are:

    a. Profiling

    b. Social media privacy threats

    c. Privacy hazards in image analytics

2.1 Digital profiling

Digital Profiling is the automated processing of person specific data to evaluate certain attributes relating to a person, particularly to analyze and predict

Figure 3.
Ethical practices in data processing.

individual's economic situation, buying habits, health, preferences, interests, behavior, etc. Digital Profiling also influences group privacy wherein an individual may be a member of one or more groups [3]. Digital Profiling is widely used in direct digital marketing businesses. Profiling without consent of the individual is a privacy breach. Cookies are a piece of data stored in the browsers, when users browse and transact. Cookies are used in auto profiling and cookies can be read without user's consent. Google has recently announced to end support for third party cookies in its Chrome browser which will make it very difficult for the digital marketing companies to build a user profile [4]. Article 22 of GDPR facilitates the right to the individual that, no automated data processing including profiling is allowed without consent from the user.

2.2 Privacy threats in social media applications

Social media platforms are highly vulnerable to stalking attacks. One of the common stalking techniques involves an online mob of anonymous self-organized groups to target individuals, causing defamation, threats of violence, and technology-based attacks. Social media are used to build trust between the perpetrator and the victim. Perpetrator is a person who may carry out a harmful, illegal and immoral act. When victim transmits confidential data including pictures and videos, the perpetrator can intercept, steal confidential data and abuses them for blackmail purposes [5]. Social media firms are also responsible to identify user with such malicious intentions, block them and initiate appropriate legal actions as per the law.

2.3 Privacy hazards in image analytics

Image data analytics is widely used in health care, social media, and e-commerce applications. In social media large numbers of images are uploaded every day. An image is worth more than a thousand words and hence it may reveal the emotional state of a person also [6]. Some of the key privacy hazards in image data analytics include

a. Attempt to analyze emotional state of people and exploit them. Facebook and WhatsApp status updates can be studied using machine learning models and sentiment analysis can help analyze the social and emotional wellbeing of a person and in turn exploit them.

b. Disclosure of secret medication being taken by a person by virtue of promotional offers on medicine.

c. Another important privacy concern is identity theft, because copies of permanent account number (PAN) cards, passports and driving licenses are kept in digital form and shared. Insurance, banking firms and third parties will extract lot of sensitive data which is a serious privacy hazard [7, 8].

d. Medical imaging deals with visual representation of internal structure of organs and tissues. Medical imaging may lead to leakage of personal and sensitive medical data of a person [9].

Ethical code of conduct in digital data processing is required for privacy preserving data processing and it is possible only through stringent implementation of privacy regulations. More than hundred countries have passed legislations to

protect individual privacy and GDPR of European Union was the pioneer. All other privacy policies and regulations are framed based on GDPR. The key features of GDPR and other privacy regulations are.

1. **Right to forget or erase data:** personal data gets uploaded in many digital applications. For example, people upload certain private photos and videos, buy certain products online and if users wish, the data or transactional records can be removed from their databases.

2. **Users consent before sharing the data:** Data holder's share and exchange data for real time insights, but in many applications the data owner is not aware of it. It is required to take the consent of the data owner before sharing.

3. **No surveillance without consent:** many applications will monitor their user's behavior including location, device type etc. Data profiling companies and digital advertisement companies do surveillance without consent from the user and most of the users are not even aware of surveillance. It is now mandatory for all firms to take user consent for surveillance, in the countries where privacy legislations are in force.

4. **Right to restrict the data processing:** many data intensive applications, process data without prior consent of the data owner. It is mandatory to take prior permission from the data owner to use data for further processing.

Failure to affirm the privacy compliance will attract serious consequences including huge amounts of fine and detriment to reputation. To ensure privacy preservation and abide by the local privacy regulations, firms are undergoing changes in their policies to incorporate privacy regulations. Following strategic changes were noticed in the year 2020 [10].

1. 30% of the businesses made changes in their cyber security models due to GDPR and 60% of them created new policies.

2. 15% firms offered extra training to staff on communications and GDPR

3. 11% firms have changed their firewalls and system configurations.

4. 6% created new contingency plans.

3. Application design strategies

Applications that cause serious privacy threats were listed in **Table 1**. For each application, design strategies and guidelines are provided, so that the applications cannot harm the privacy of the user.

3.1 Design guidelines for smart phone apps

It is a common practice that most of the users do not read the privacy policy and the network permissions which an app demands before installation. People ignore and will agree for all permissions the app demands which lead to serious privacy concerns. To ensure inherent privacy protection, smart phone apps must be designed with following features.

a. Seek only the minimum permissions for the app to be functional.

b. Do not collect any metadata including location, type of device, time etc.

c. No auto profiling of the user by any app is allowed.

d. Accept and abide the federal laws of the region or state pertaining to data access and sharing.

e. Design to ensure no access to any free Wi-Fi which is not registered by the user.

f. Do not transfer any data from the phone without consent from the user.

g. Privacy policy should not be a text document. Privacy policy should be an audio file played in the language opted by the user, ensure the user listens to it completely and finally accepts or rejects the privacy policy. Polling can be used to ensure user's attention. i18n (internationalization) applications are required and easy to develop with present open source technology frameworks to offer privacy policy as an audio file in the language opted by the user. i18n applications are the applications that offer multilingual user interface. It is the process of writing software so that it can support local languages and cultural settings.

i18n applications: The word i18n represents internationalization. In the word "internationalization", the number of characters between the first and last characters i.e. i and n are 18, hence the name i18n. Applications are said to be i18n applications when they support multilingual user interface. Applications read the request headers to know the language preferences of the user. For example if the user's language preference is Spanish, then the user interface will automatically reflect the content in Spanish. Generally i18n is applied to web applications. In web applications, http protocol is used for request and responses. When a http request is made, along with the request few request headers will also be sent to the web application. One of the request header is "*accept-language*" which contains the language which the user prefers to use. These language preferences can be changed by the users through browser settings. If the web application is i18n enabled then it will read the value of the *accept-language* header and display the user interface in the language mentioned by the user. Such applications are called i18n applications.

3.2 Design guidelines for e-commerce sites

e-commerce sites use recommendations to offer value added services to the customers. Recommendations are used as part of improved service. However, there is always a possibility of information disclosure. For example, a person wanted to buy some product for personal use. He/she wanted this to be confidential and by virtue of recommendations, he/she may see a pop up or alert showing a better offer on that product which is visible to the people sitting nearby and this will lead to discrimination and personal embarrassment. Based on the type of products bought, the gender of the person can also be inferred which is an unwanted disclosure. In order to ensure privacy protection, following features need to be incorporated in the design of the e-commerce sites in line with the privacy legislations.

1. Privacy Quotient (P_μ): Recommendations are used by ecommerce firms to provide value added services and best possible offers to the customers based on their buying habits and transaction history. Recommendation systems lead to

serious privacy concern which is not addressed by any ecommerce firm and the same is illustrated here. For example a person regularly bought some product online, related to personal care and does not want to disclose this to anyone. However, since it is a regular transaction the ecommerce firm would like to recommend the same product to him by offering decent discount on the product and the same will displayed on his screen when he/she logs into their account and it is a privacy breach if someone else sees the same. It can lead to discrimination of the person in the family or profession. To address this problem, we introduce the concept of privacy quotient. For every product the ecommerce firm should provide an option where in user can opt, whether this product and purchase is to be made private or not, thereby excluding it from any form of analytics or recommendations. If 40% buyers of a product opt for transaction privacy i.e. the product purchase is not to be used for recommendations, then the product must be considered as private and for all buyers of this product, the transaction must be made private. This percentage of transactions which decide the transaction privacy is called as privacy quotient (P_μ) [11].

2. No sharing of data without users' consent: No e-commerce site, must share customers data without consent. However, data can be shared with federal authorities for any investigation purpose.

3. Meta data: e-commerce sites tend to collect metadata including location, type of device used, IP address etc. without the permission and knowledge of the user. It has to be avoided.

3.3 Social media platform design issues

Social media has emerged as the most vulnerable platform of privacy abuse especially cyber stalking, ransom ware, sexual abuse etc. Important issues to be addressed in social media applications are

1. Identification of fake accounts and stringent mechanism of anomaly detection.

2. Deep neural networks can be used in identifying the private and sensitive information in the images uploaded by the user, remove them and store the modified image. User consent is mandatory. Users must be advised of privacy threats every time when they upload photos or videos.

3.4 Data capturing systems

Disclosure and discrimination are the common threats related to data capturing systems. Organizations like hospitals, banks, retail supply chain etc. collect a lot of person specific data while offering respective services. This data will be analyzed to gain deep insights and come up with better decisions and offer value added services.

1. As per the privacy regulations across many countries, it is recommended to use non-anonymized and model based solutions for privacy preservation.

2. Sensitive attributes must be tokenized before sharing with any other third party for analytics.

3. Quasi identifiers must be synthesized before sharing.

3.5 Data privacy officer

Every organization that deals with personal and sensitive data must employ a Data Protection Officer (DPO). DPO must be a technology expert with sound knowledge on privacy policies and regulations.

DPO is responsible for ethical code of conduct and implementation of privacy laws of the respective region or territory. Some of the key responsibilities of the DPO are

1. Provides complete security to the data.

2. Records all the activities performed on the data.

3. Seeks consent from the data owner, every time when the data is processed

4. Responds to the queries of the customers or data owners.

5. Ensures implementation of local privacy policies and federal laws.

6. Notification of privacy breach if any to the data owner

7. Impact assessment

4. Conclusions

As part of our work, we proposed few guidelines for application design which will support individual privacy in many data intensive applications in line with privacy legislations. These days more privacy violations and abuse are being reported in social media where people upload lot of personal photos and videos. Huge number of fake profiles were also reported who may indulge in activities like cyber stalking, ransom ware etc. There is a need for strong and coherent privacy preservation mechanism for social media applications and has enough scope for research especially employing deep learning models.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author details

Ram Mohan Rao P[1*], S. Murali Krishna[2] and AP Siva Kumar[1]

1 Department of Computer Science and Engineering, JNTUA Ananthapuramu, Andhra Pradesh, India

2 Department of Computer Science and Engineering, Sri Venkateswara College of Engineering Tirupathi, Andhra Pradesh, India

*Address all correspondence to: rammohan04@gmail.com

IntechOpen

References

[1] Ducange, Pietro, Riccardo Pecori, and Paolo Mezzina. "A glimpse on big data analytics in the framework of marketing strategies." Soft Computing 22.1 (2018): 325-342.

[2] Dhillon, Gurpreet, and Kane J. Smith. "Defining objectives for preventing cyberstalking." Journal of Business Ethics 157.1 (2019): 137-158.

[3] Mavriki, Paola, and Maria Karyda. "Automated data-driven profiling: threats for group privacy." Information & Computer Security (2019). https://doi.org/10.1108/ICS-04-2019-0048

[4] https://www.cnbc.com/2020/01/14/google-chrome-to-end-support-for-third-party-cookies-within-two-years.html

[5] March, Evita, et al. "Somebody that I (used to) know: Gender and dimensions of dark personality traits as predictors of intimate partner cyberstalking." Personality and Individual Differences 163 (2020): 110084. https://doi.org/10.1016/j.paid.2020.110084

[6] Chen, Lushi, et al. "Building a profile of subjective well-being for social media users." PloS one 12.11 (2017). https://doi.org/10.1371/journal.pone.0187278

[7] Yang, Jingjing, Jinzhao Wu, and Xiaojing Wang. "Convolutional neural network based on differential privacy in exponential attenuation mode for image classification." IET Image Processing (2020). https://doi.org/10.1049/iet-ipr.2020.0078

[8] Beaulieu-Jones, Brett K., et al. "Privacy-preserving generative deep neural networks support clinical data sharing." Circulation: Cardiovascular Quality and Outcomes 12.7 (2019): e005122. https://doi.org/10.1161/CIRCOUTCOMES.118.005122

[9] P. Wang, T. Chen and Z. Wang, "Research on privacy preserving data mining," Journal of Information Hiding and Privacy Protection, vol. 1, no.2, pp. 61-68, 2019.doi:10.32604/jihpp.2019.05943;

[10] https://www.nvtgroup.co.uk/keep-safe-this-data-privacy-day/

[11] Rao, P. Ram Mohan, S. Murali Krishna, and AP Siva Kumar. "Novel algorithm for efficient privacy preservation in data analytics." (2021).