Why pay attention to the data security of digital media platform in the context of big data analysis ?

Jiayi Cui

Computer Science School, University of New South Wales, 2052 NSW, Australia

cjy525792926@gmail.com

Abstract. Data security has been a topic of widespread concern. When people surf the Internet, they tend to push the same content on multiple websites. If they view a certain type of video multiple times, they will be frequently pushed similar content, and even their purchases will be "monitored." The development of big data has brought convenience to People's Daily life, but with the penetration of big data into daily life, many behaviors in daily life has also been recorded by big data. Once the information of the database is leaked and used by criminals, irreversible losses will be caused. This paper analyzes and summarizes the phenomenon of digital media platform data leakage, and multi-dimensional data protection and further guarantees data security in the future, which can better help solve the data security problems brought by the development of digital media platforms.

Keywords: Digital media; Data security; FPE and encryption SECUREDL system; Data anonymization protection techniques

1 Introduction

With the advent of the Internet era, more and more digital media platforms utilizing big data emerge. Among them, the social platform is the focus of attention. Social platforms have a wide range of functions, and there are some social functions that can realize functions as well as entertainment and shopping, but there are also some data security problems. Whether social platforms have a way to ensure that social content is not stolen, that personal information is kept in the database when the corresponding content is pushed, and that it is not used by criminals are all issues that need to be discussed.

2 Literature Review

2.1 Why to use big data technology in digital media platforms?

According to relevant statistics, global digital media users have reached more than 50% of the world's population, and every other person on average uses digital media in their daily life. Digital media platforms include social platforms that combine traditional

H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 213–218, 2022.
https://doi.org/10.2991/978-94-6463-108-1_24

social activities with other functions, such as shopping and viewing interesting videos. This requires the use of big data technology to analyze personal preferences and hobbies and requires users to provide payment information and collect a large amount of data for analysis. Although this part needs to use algorithms, it also needs big data technology to assist the analysis of the benefits to data, while some problems arise.

2.2 Why to study data security of digital media platforms:

With the development of the Internet and the advent of the digital era, digital media platforms are developing in accordance with the trend of the times. The current digital media platforms are no longer just a single platform, but a life platform with multiple functions, convenience and entertainment. The leakage of data information is not only the leakage of basic information such as account numbers and passwords. Most platforms require real-name authentication, including the acquisition of identity information. Some platforms also have databases that store user addresses and other information, which can be used by criminals once leaked. Other fields, such as finance and healthcare, will face the same problems as The Times go on. However, there are summaries of research **[5]** on the leakage of medical and health databases **[4]**, information leakage in the financial field and data leakage in other fields, but there are not many summaries on the information leakage of digital media platforms. This paper will analyze and summarize the harm and solutions to this problem.

2.3 Traditional approaches to data security

Big data is a field that has been growing rapidly in recent years. It refers to the large amount of data produced at an exponential rate with the development of the Internet. N. Miloslavskaya **[2]** pointed out that big data is not only stored in its raw form as before but also in data pools in structured or unstructured form. It is mentioned in **[1]** that hackers attack databases for their sense of achievement. While authorities have been aware of this and improved data security, hackers are also constantly improving their techniques to find vulnerabilities in programs. Therefore, the data security of digital media has always been a problem of great concern and urgent need to be solved, but simple data security technology is not enough to deal with the current situation, so more complex and more difficult to crack methods are needed.

3 Harming of Data leakage on new media platforms

In the past few years, there have been a number of serious data breaches on digital media platforms. In 2017, amazon's S3 cloud storage database accidentally leaked the personal information of 1.8 billion people due to a configuration error. In 2013, because of yahoo's poor data security, hackers stole 3 billion account information, causing huge losses to the company. In addition to these events, incidents of information leakage are still frequent. With the development of Internet technology, People's Daily life has been connected to the Internet. The user information of digital media platforms not

only includes the user's basic information (real-name authentication, password, real-name, identity information) but also may involve the user's address and other information in some payment platforms with shopping functions. At the same time, users are accustomed to using the same password across multiple platforms, which can cause property damage. With the continuous development of digital media platforms, the more functions it has, the more convenient it is for users, but it means that there will be more and more data security risks. This needs to improve data security should also change hackers and fraud, but also needs the joint efforts of the whole society to protect the full range of personal information awareness.

4 The Method of Data Security for new media Platforms

In the previous section, we examined the dangers of data leakage. In this section, we will discuss how to address these dangers.

4.1 FPE and encryption

Data leakage is the biggest threat to big data, and encryption can be used as the main protection tool **[1]**. However, data needs to be decrypted during transmission, which poses a threat to data security. To address this problem, we explored a data masking scheme using Format retention encryption (FPE) for large amounts of data using the Spark framework in [3]. Although this method requires the algorithm to call the block cipher many times, which is not efficient, the method chooses different FPE algorithms according to the data type and requirements, and the ciphertext still retains the original plaintext format and does not contain any sensitive information and is not an unreadable binary string.

4.2 Data anonymization protection techniques

Data anonymization **[6]** is about protecting data by eliminating indicators. In this process, the user's data source is hidden to protect the data, but an attacker can also use de-anonymization techniques to extract information. Therefore, to prevent such attacks, relevant departments must protect their data from unauthorized access and comply with relevant privacy regulations.

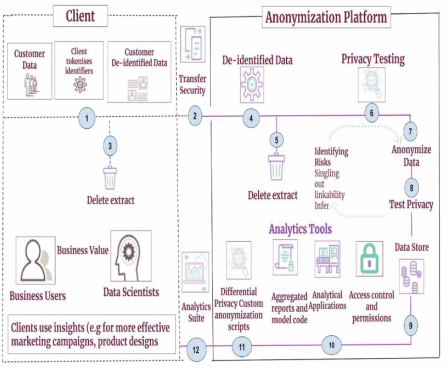

Fig. 1. Data is anonymized

4.3 SECUREDL system

4.3.1. Summary of SECUREDL system.

Specifically, the SECUREDL system is speculative and allows the following aspects to be organized. 1) Formulate policies for sensitive data access. 2) Automatically maintain necessary audit logs in compliance with laws and regulations. 3) clean and revise sensitive data in real-time based on data sensitivity and AI model requirements. 4) detect potential unauthorized or abnormal access to sensitive data. 5) Automatically create attribute-based access control policies based on data sensitivity and data type. Compared to some existing approaches, SECUREDL provides multiple capabilities across multiple data management systems, including data cleansing, detection, intrusion detection, and data management.

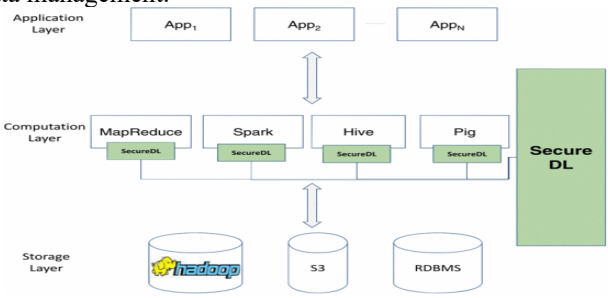

Fig. 2. SECUREDL system

4.3.2 Architecture description.

As shown in Fig. 2, SECUREDL is a data access agent designed to be built on Hadoop or Spark. SECUREDL captures information submitted by users and analyzes and processes it according to security and privacy policies. At the same time, using machine learning models built from access history, Securedl can determine if the current request is materially different from past requests. This can be used to detect potential network

attacks, including abuse by insiders, if the network attack has a data access pattern that is significantly different from regular data access requests. The idea is to protect data in several ways, but this approach has some limitations. If the user only uses our system and the provider has no other way to attack the data, the attacker may try to place malicious code to circumvent the injection strategy. To address this challenge, there is a great deal of research in the literature (e.g. **[8]**) to limit untrusted code and programs. In our example, we leverage these types of existing technologies (for example, Java Security Sandbox) to prevent any data access other than the data access interface specified by the underlying NoSQL database. In addition, restrictions will be set at the Java virtual machine level to prevent submitted jobs from using other potentially malicious libraries, such as reflection.

5 Challenge and conclusion

Data security is a topic derived from the development of big data. As we continue to develop ways to improve data security, our technology for hackers is also constantly updated. In order to better protect the data security of digital media platform, we should not only constantly improve the technical level and thinking, but also make efforts in many aspects. Information security technical personnel actively study methods to protect data security at the same time, the public should also enhance the awareness of information protection and the state issued relevant data security regulations and policies and publicity. Only with concerted efforts can the problem be fundamentally solved. How to do it effectively and efficiently will also be a big challenge in the future. There is a general awareness that technological improvement is a long process and will take time. Even if you know the way, you can't solve the problem overnight and for good.

This article on digital media platform has carried on the simple analysis summary, from why use big data technology, the data security problem brought by the large amount of data, and finally summarizes some about data security solution in literature, hope it'll be helpful for future related research.

References

1. D. B. Rawat, R. Doku and M. Garuba, "Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security," in IEEE Transactions on Services Computing, vol. 14, no. 6, pp. 2055-2072, 1 Nov.-Dec. 2021, doi: 10.1109 / TSC. 2019.2907247.
2. N. Miloslavskaya and A. Tolstoy, "Application of Big Data, Fast Data, and Data Lake Concepts to Information Security Issues," 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), 2016, pp. 148-153, Doi: 10.1109 / W - FiCloud. 2016.41.
3. B. Cui, B. Zhang and K. Wang, "A Data Masking Scheme for Sensitive Big Data Based on Format-Preserving Encryption," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2017, pp.518-524, doi: 10.1109/ CSE-EUc.2017.97.

4. H. Kupwade Patil and R. Seshadri, "Big Data Security and Privacy Issues in Healthcare," 2014 IEEE International meets on Big Data, 2014, pp. 762-765, doi: 10.1109 / BigData meets. 2014.112.
5. D. Feng, W. Hao and T. Ding, "Network Finance Security Problems and Countermeasures," 2009 International Symposium on Computer Network and Multimedia Technology, 2009, pp. 1-4, doi: 10.1109 / CNMT. 2009.5374654.
6. S. Varshney, D. Munjal, O. Bhattacharya, S. Saboo and N. Aggarwal, "Big Data Privacy Breach Prevention Strategies," 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 2020, pp. 1-6, Doi: 10.1109 / iSSSC50941.2020.9358878.
7. M. Kantarcioglu and F. Shaon, "Securing Big Data in the Age of AI," 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), 2019, pp. 218-220, doi: 10.1109 / TPS - ISA48467.2019.00035.
8. Mengtao Sun, Gang Tan, Joseph Siefers, Bin Zeng, and Greg Morrisett. 2013. Bringing java's wild native world under control. ACM Trans. Inf. Syst. Secur. 16, 3, Article 9 (November 2013), 28 pages. https://doi.org/10.1145/2535505