International Journal of Electrical, Electronics and Computer Engineering **13**(1&2): 05-06(2024)

ISSN No. (Online): 2277-2626

Privacy-Preserving Techniques in Web Mining

Akshat Sharma*

Department of School of Computer Science and Engineering
Govt. P.G College Dharamshala, Himachal Pradesh Technical University (HPTU) (Himachal Pradesh), India.

(Corresponding author: Akshat Sharma)*

(Received: 04 January 2024, Accepted: 15 March 2024)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Web mining has now become the most important technique for getting valuable information from a large amount of data. However, as the reliance on web mining grows, concerns about privacy increase. This research paper aims to provide privacy prevention techniques in web mining, focusing on methods such as anonymization, encryption, and differential privacy. This paper will also help us to explore the importance of privacy and security with mining and also legal and ethical concerns about web mining.

Keywords: Web mining, data breaches, security.

INTRODUCTION

Web mining is a process of extracting useful information from the World Wide Web. This process involves extracting useful patterns and knowledge from the web. With the significance of web mining in various domains, some security concerns are becoming paramount. The main aim of this paper is to explore privacy-preserving techniques to secure data, by analyzing some cases (Kosala, 2000; Kumar, 2015). Several risks and challenges are associated with security and privacy in the realm of web mining:

1. Unauthorized Access
2. Data Breaches
3. Misuse of Sensitive Information
4. Lack of Anonymity
5. Ethical Concerns
6. Adherence to Privacy Regulations

In this paper, we will delve into privacy-preserving techniques that address these risks. By examining relevant case studies, we aim to provide insights into effective strategies for securing web-mined data and mitigating potential privacy breaches. The objective is to contribute to the ongoing discourse on maintaining the delicate balance between extracting valuable knowledge from the web and safeguarding the privacy of individuals and organizations. (LinkedIn) (Vipula Vinaykumar Mahindrakar*, oct 2018)

METHODOLOGY

Here by analysing some cases, we are going to conclude major privacy protection techniques that are lacking in these companies.

Case Studies:

Yahoo Data Breach (2013-2014): The Yahoo data breach, which occurred in 2013 is considered one of the most significant breaches in history which compromises almost 3 billion users' accounts. This event is disclosed in late 2016. The cybercriminals of this breach are intended to be state-sponsored hackers that received unauthorized access to the Yahoo security systems to steal a significant amount of customers' information, like their names, phone numbers, email addresses, birth dates, hashed passwords, and many more. This incident was undetected for many years which highlights a lapse in Yahoo's security features and incident reaction capabilities. Yahoo faced a lot of criticism for disclosing this event because the event was disclosed in 2016. This raises concerns about the company's transparency and data protection abilities. This incident increases the user's concern about the safety of their data. This also leads to a negative impact on the company and also compromises the user's trust. This incident increases the awareness of cybersecurity and also increases the awareness of organizations to prioritize cybersecurity measures to make their data secure. After this incident, yahoo took some steps to enhance its security measures by implementing encryption methods and also by improving its incident response procedure. This incident also serves as a lesson for other organizations to increase proactive security measures, timely disclosure of attacks, and continuous improvement in the face of increasing cyber threats (Wikipedia, 2013; Redfern, 2013; Online, 2020).

Equifax Data Breach (2017):

The Equifax data breach in 2017, was one of the significant cybersecurity incidents that affected one of the largest credit reporting agencies. In this breach, the sensitive data of approximately 147 million users was affected. The cybercriminals exploit a vulnerability in Equifax's website to get unauthorized access to large amounts of sensitive data including names, social security numbers, birthdates, addresses, etc. This breach increases the risk of identity theft and the stolen data could be used for various other activities. The major fact of this breach was the nature of the information exposed, this information looks like it is theft for identity theft. From this, the company faced a lot of criticism for its cybersecurity practices and also for the late discovery of the breach. After this incident, there is a sudden increase in concerns about the security of personal information and also the need for strong data protection measures. This incident also questions the credit reporting companies about safeguarding the user's sensitive data. To overcome the effect of this breach Equifax took some steps such as free credit score monitoring and also made some security improvements to prevent similar incidents in the future. This case serves as an important point to understand the need for robust security measures, mainly for the companies that are holding sensitive data (Miyashiro, 2021) (KKU, 2024) (CSO, 2020).

RESULT

Through the analysis of these cases, there are many lacks that companies were facing at that time. So based on these case studies the major prevention techniques in web mining to make the data secure are:

1. Data Encryption: It is important to make the data encrypted at the time of transmission and storage. This process helps to prevent the information and password at the time of breach.

2. Regular Security Audits and Assessments: This majorly helps to check the system's vulnerabilities. So we can regularly check the system's weaknesses,and prevent them from being exploited at the time of cyber attack.

3. Transparency and Communication: Transparency in communication not only helps to rebuild trust but also showcases the commitment to resolve these issues promptly.

4. User Authentication Measures: Strengthening user authentication enhances account security. By adopting the multifactor authentication an additional layer is included which reduces the unauthorised access.

5. Employee Training: To reduce the errors by humans in most of the breaches companies have to educate employees on cybersecurity, threats, and the importance of safeguarding sensitive information.

6. Third-Party Vendor Security: Most of the breaches occur from third-party systems. So to overcome this risk companies have to scrutinize the vendor's security practices.

By distilling insights from historical breaches, organizations can extract valuable lessons to fortify their privacy measures. Implementing a combination of encryption, regular assessments, transparent communication, robust authentication processes, regulatory compliance, employee training, and rigorous third-party vendor scrutiny collectively contributes to a more resilient and secure data environment**.**

CONCLUSION

This comprehensive exploration of web mining, encompassing its methodologies, security concerns, privacy-preserving techniques, case studies, and future trends, illuminates the multifaceted landscape of extracting insights from the World Wide Web. Key findings and insights emerge from each fact, culminating in a holistic understanding of the challenges and opportunities inherent in this dynamic field. The importance of security and privacy in web mining cannot be overstated. The potential risks, including unauthorized access and data breaches, necessitate robust measures to safeguard sensitive information. Ethical considerations are paramount, urging practitioners to strike a delicate balance between deriving insights and respecting user privacy. Anonymization, encryption, and differential privacy emerge as powerful tools in the arsenal of privacy-preserving techniques. These methodologies ensure the confidentiality and integrity of user data, addressing ethical concerns and regulatory requirements. The real-world incidents of the Yahoo Data Breach (2013-2014) and the Equifax Data Breach (2017) serve as cautionary tales. These breaches, with their far-reaching consequences, emphasize the need for robust security practices, transparent communication, and adherence to legal and ethical standards.

REFERENCES

CSO (2020). Retrieved from CSOONLINE: https://www.csoonline.com/article/567833/equifax-data-breach-faq-what-happened-who-was-affected-what-was-the-impact.html

KKU (2024). EQUIFAX DATA BREACH. Retrieved from WIKIPEDIA.

Kosala, R. (2000). web mining research: a survey.

Kumar, S. N. (2015). World toward web mining. american general of system and software.

Linkedin. (n.d.). How can you protect your data privacy and security when web mining. linkedin.

Miyashiro, I. K. (2021). Case Study: Equifax Data Breach.

Online, B. (2020). Yahoo Data Breach: What Actually Happened ? BPB.

Redfern, B. E. (2013). The Yahoo Cyber Attack & What should you learn from it ? cashfloat.

Vipula Vinaykumar Mahindrakar, B. L. (2018). Web Mining Opportunities and Privacy Challenges. Ignited Minds Journals.

wikipedia. (2017). yahoo data breach.