COMPUTER SCIENCE AND DATA ANALYTICS GUIDED RESEARCH I

Midterm Report

Student: Gunay Maharramli

Title: Data Privacy and Security in Database System

Instructors: Prof. Dr. Stephen Kaisler, Assoc. Prof Dr. Jamaladdin Hasano

Problem Description

As organizations increasingly rely on database systems, the security of the information they handle becomes crucial. Data security and privacy are built on the principles of Confidentiality, Integrity, and Availability (CIA), but there are concerns about whether modern database systems can effectively uphold these principles in their architectures. Moving database infrastructures from on-premise to cloud-based setups has also raised the risk of security and privacy breaches. As a result, many organizations prefer not to store their most critical data in the cloud, as they believe on-site storage provides a higher level of security confidence. Hence, this research paper places its emphasis on the security and privacy aspects of a PostgreSQL database, prioritizing measures that ensure a secure and private environment for the database system. The study focuses on existing security models in PostgreSQL database system and explores ongoing research efforts to strengthen and improve these security mechanisms. PostgreSQL is a popular open-source relational database management system known for its strong security features and reliability. My selection of PostgreSQL is guided by several reasons. Firstly, it comes with built-in security mechanisms, such as robust authentication methods and access controls, which help protect data. Secondly, PostgreSQL places a high priority on data integrity, ensuring that any changes made to the data are reliable and consistent. Another reason for choosing PostgreSQL is its active and supportive community. This community continually works on improving the database's privacy and security features, promptly addressing any issues that arise and providing regular updates. In my research, I will focus on analyzing the security aspects of PostgreSQL, including its access controls, encryption capabilities, and auditing mechanisms. The ultimate goal of this research is to provide insights into enhancing database security, benefiting both professionals and researchers in the field.

Strategy Definition:

In this research paper, the chosen strategy for data collection revolves around qualitative and quantitive data.

Using qualitative data for this study allows for a more nuanced exploration of the security and privacy aspects of PostgreSQL database systems. It enables a deeper understanding of the

perspectives, concerns, and challenges faced by users, developers, and administrators in safeguarding data within these systems. Through methods like interviews, focus groups, and observations, the research can capture the complexities and contextual nuances that quantitative data might overlook. By employing qualitative data, this research paper seeks to provide a comprehensive and holistic view of the security and privacy landscape of PostgreSQL databases. It will enable the exploration of real-life scenarios, uncovering potential vulnerabilities, identifying best practices, and understanding the underlying reasons for certain security and privacy practices. This approach will contribute to a more nuanced and well-rounded analysis, helping to inform practical strategies for enhancing database security and privacy while also providing valuable insights for future research in this domain.

The quantitative method used in this research involves assessing the number and types of vulnerabilities discovered in popular database system over a specific period, typically over time. By systematically collecting data on identified vulnerabilities in these databases, researchers can quantify and analyze the trends and patterns in security weaknesses. This approach aims to provide empirical evidence regarding the prevalence and nature of vulnerabilities, which can offer valuable insights into potential weaknesses that require prompt attention and mitigation. The figure depicts the number of vulnerabilities identified in the PostgreSQL database over a specific period, providing valuable insights into its security weaknesses.

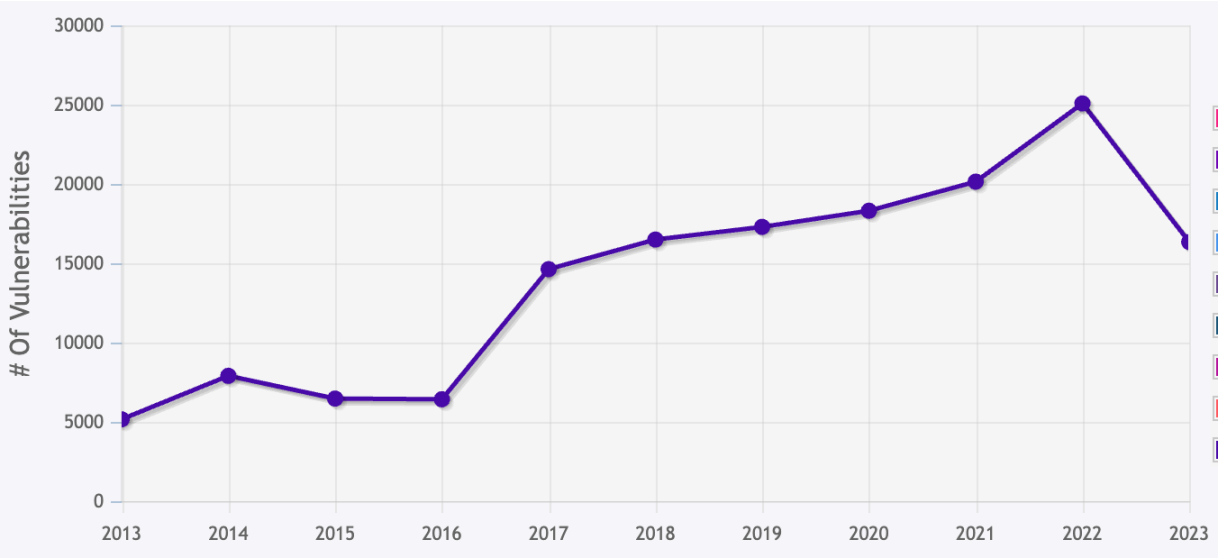

Figure 1. Number of Vulnerabilities Over the Time based on CVE details.

Progress Overview:

In this research, the investigation of database privacy and security at the database level, with a focus on PostgreSQL, involved a series of tasks that provided a comprehensive understanding of the subject matter. Research design was formulated, outlining the objectives, research questions, and methodology that guided the study. Extensive reading of scholarly articles related to database privacy and security was conducted. The methodologies, findings, and recommendations put forth by experts in the field were critically evaluated and analyzed.

I. Transaction Models

The concept of transactions and their logical meaning has developed alongside data management techniques. Managing concurrent transactions becomes a concern when multiple users access the same dataset in a database, as it requires ensuring data consistency and integrity. The CAP theorem, also known as the Brewer's theorem, states that it is possible to achieve, at most, two out of the three properties - Consistency, Availability, and Partition tolerance - in a distributed computer system.

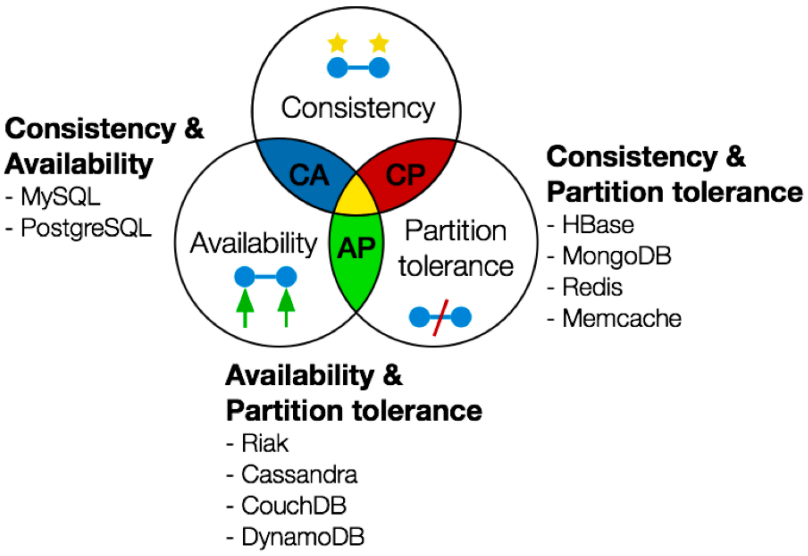

Figure 2. Database Systems and CAP Theorem.

II. Database System Attacks

In general, attacks on database systems can be classified into two main categories: the first involves compromising data confidentiality, while the second pertains to the revelation of data privacy.

- Attacks based on Confidentiality of Data that includes Injection Attacks, Inference and Reconstruction Attacks, Concrete Attacks, Snapshot Leaks, Full System Compromise
- Attacks based on Privacy of Data: Correlation Attacks, Identification Attacks

III. Protection Mechanisms

- Authentication is a process that validates and confirms the identity of users attempting to access a database system before granting them permission to utilize its data and resources. This can be implemented in various ways, spanning from individual user authentication to mutual authentication, ensuring both the user and the database server confirm each other's identities before establishing a connection. Currently, the majority of relational database systems come equipped with authentication mechanisms in place.
- Authorization is a crucial aspect of the security framework in any database system. After confirming the user's identity through authentication, the next step is to associate and permit the user access to specific resources within the database. This process, known as authorization, ensures that only authorized users or roles are granted access to predefined sets of objects or the entire database.

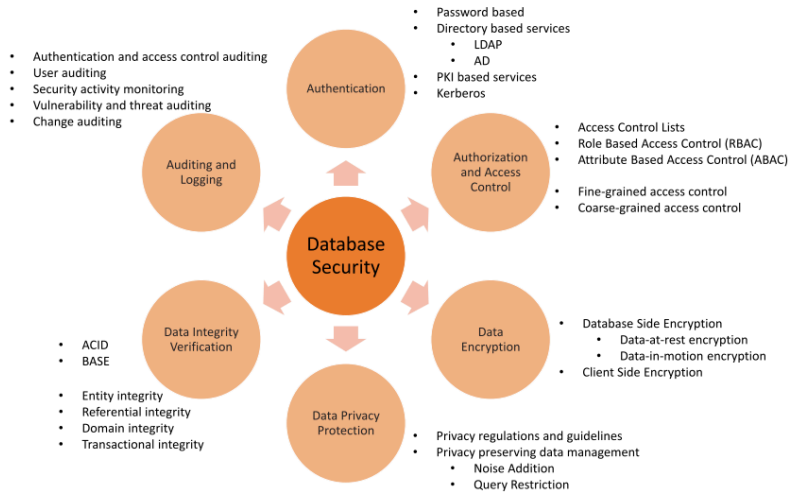

Figure 4. Database Security Mechanisms

IV. Conclusion

Through extensive research and data gathering, this study has shed light on various possible attacks on database systems. The gathered information has highlighted the critical need for robust security measures to safeguard databases from potential threats. By understanding the vulnerabilities and attack techniques, organizations can better prepare and implement effective security strategies to protect sensitive data and ensure the integrity and confidentiality of their databases. For this project, an experimental setup was established using a PostgreSQL database. The database was carefully configured to implement appropriate security measures, such as access controls, auditing mechanisms. Security measures, including access controls were implemented and evaluated within the PostgreSQL database. The effectiveness of these measures in protecting sensitive data and their impact on performance were assessed.

Database	Authentication	Authorization	Consistency Model	Auditing and Logging
PostgreSQL	Different types of Authentication available like LDAP, SSPI	Role based permissions	ACID	Different ways exists