Confidentiality Preservation in Big Data Centres

Proposal

Supervisor:
Jason Jaskolka

Team Members:
Aleksandar Savic - 100999110
Hasan Issa - 100921446
Ryan Zheng - 100996797
Calvin Soong - 100999332
Tashfiq Akhand - 101004428

Table of Contents

Background

Users/customers across the globe expect their privacy to be taken seriously, thus making it a responsibility for companies around the world to keep their databases secure and the data stored within from being compromised. Companies that have an online presence must consider database security as a priority, but this task becomes much more difficult as data centers grow. We have seen many publicly listed companies and banks report data breaches over the course of the last few years which has had serious ramifications for them both in terms of finances and reputation.

An example of a database breach would be 2 powerhouse banks such as BMO and CIBC. The victims of the banks were demanded to send a million dollars in cryptocurrency from hackers [1]. The danger of the matter were names, social insurance numbers and security questionnaire answers of the customers which are instrumental information towards personal corruption [1]. The attackers were able to break and breach both banks database security defenses and had attacked tens of thousands of clients [1]. This demonstrates enlarged chaotic damages in terms of financial and legal logistics when a data breach occurs. Many causes are because the database walls aren't fully protected by inference attacks.

Objectives

The objectives of this project is to identify and generate a suitable application and obtain a suitable case study data set in order to stop attackers from getting and piecing together information at one security level to determine a fact that should be protected at a higher security level. The team will develop a mechanism to specify the system confidentiality policy and implement mechanisms and software tools for identifying potential inference attacks violating the specified confidentiality policy. The implemented mechanism will also deploy strategies to enforce the specified confidentiality policy for the user. The system will keep a log of previous user queries in order to properly assess safe results for future queries. The system developed will be suited for tabular and relational format databases.

The system will not be handling colluding users attempting to perform a mass inference attack, the system will assume it is one user attempting to perform the inference attack . The system will assume that the database has already been built.

Methods

This project will help create a system that ensures the confidentiality policy of a data center remains intact even through authorized queries. Clients of the system to be developed will dynamically be able to input security protocols/rules into the system to protect their sensitive data from being accessed by unauthorized individuals through inference attacks.

The system will keep a log of all the queries made by a user and using an algorithm, will be able to identify if any sensitive data can be deduced through previous queries. If any vulnerabilities are detected, the system can respond using:

PolyInstantiation:

PolyInstantiation is a method that allows information in a database to have different levels of confidentiality and can be acquired depending on the level of the authorized user that is requesting access [2]. This method is an SQL related terminology that allows rows of multiple quantities to have the same primary key but, the instances of the rows are determined by the authorized security level [2]. This allows the protection of sensitive client data/information which protects a database from many inference attacks since the wanted resources of the attackers are in a hierarchy leveled security.

Cell Suppression:

Cell suppression is a method that can be used to protect sensitive data being queried from the database. Cell suppression hides a combination of rows and columns of data from the querying user to prevent the user from inferring any sensitive relationships between the data [3]. This can be performed when a possible inference attack is detected, and the data queried by the user can be modified using rounding or by generating random values. Rounding is a method that multiplies all values in the table/column to a multiple of the rounding base [3].

Task Breakdown

Aleksandar	● Responsible for front-end development
Ryan	● Responsible for front-end development ● Responsible for configuring the database
Calvin	● Responsible for developing the web-server ● Responsible for Back-end development
Tashfiq	● Responsible for working on the algorithm in the back end used to enforce the confidentiality policies
Hasan	● Responsible for working on the algorithm in the back end used to enforce the confidentiality policies

Timetable

Deliverable	Due Date	Remarks
Project Proposal	**Monday, September 30, 2019**	Submitted Online by Noon
Requirements Specifications	Tuesday, October 15th, 2019	
High-Level Design	Wednesday, October 30th, 2019	
Detailed Design	Wednesday, November 20, 2019	
Progress Report	**Friday, December 6, 2019**	Submitted to Supervisor
Implementation & Prototype	Tuesday, January 14th, 2020	
Oral Presentation	**Week of January 27, 2020**	Submit Oral Presentation Form online by Friday, December 6, 2019
Testing (V&V) Documentation	Wednesday, February 12th, 2020	
User Documentation	Wednesday, February 26th, 2020	
Final Report Draft	**Monday, March 9, 2020**	Submitted to Supervisor
Poster Fair	**Monday, March 16, 2020**	Submit Poster Fair Demo Form online by Monday, January 6, 2020
Video Demonstration	**Competition Deadline**	Stay tuned to Course Webpage
Final Report	**Tuesday, April 7, 2020**	Submitted Online and to Supervisor

Benefits

Attackers have developed many tactics to conduct direct attacks on databases including SQL injection and privilege abuses. A much more subtle attack that has been used in the past is inference attacks. An inference attack is a process by which an attacker performs authorized queries and deduces (through reasoning) unauthorized information from the legitimate responses received therefore resulting in a violation of the data centre's confidentiality policy [4]. Inference attacks highlight the importance of the design of a big data centre and truly show the benefits of this project.

Larger scale companies with huge data centres will have difficulties identifying how and through what queries their sensitive data can be breached through inference attacks. Also, if the design of their databases is shown to be flawed against any inference attacks, the changes to the infrastructure can prove to be costly and time-consuming.

This project will help clients by saving them time and money. The system to be developed will allow the client to dynamically change the security protocols of his/her data center. This will prove very time and cost friendly to the client by eliminating the need to make any big modifications to the design and infrastructure of the database. The data centre's confidentiality policy is enforced simply through the clients inputs into the system.

Having to search through vast databases and identify any possible inference attacks on sensitive data is very exhausting and time consuming. The system to be developed will save the client's time by automating this process and eliminating any possible inference attacks to be made on sensitive data. Using a self-proprietary algorithm, the system will be able to detect if and when any sensitive data can be deduced through previous queries made by the user and will then respond accordingly.

References

[1] Solomon, H. (2019). *BMO, CIBC victims of cyber breach, attackers demand $1 million from each in cryptocurrency*. [online] IT World Canada. https://www.itworldcanada.com/article/bmo-cibc-victims-of-cyber-breach-attackers-demand-1-million-from-each-in-cryptocurrency/405703 [Accessed 23 Sep. 2019]

[2] Skillset.com. (2019). *What is the 'main' purpose of polyinstantiation? - Skillset*. [online] https://www.skillset.com/questions/what-is-the-main-purpose-of-polyinstantiation

[3] Turkanovic, M. Družovec, T. Hölbl, M. (2015) *Inference Attacks and Control on Database Structures*

[4] Cybrary. (2018). *Inference Attack*. [online] Cybrary.IT. https://www.cybrary.it/glossary/i-the-glossary/inference-attack/

Student Paragraphs

Calvin -

This project is suitable for my 4th year project because it emcompases the key components of Software Engineering in it. My knowledge gained throughout my undergraduate years can be applied to the many tasks in this project, such as requirements gathering, software design, software implementation, testing, and documentation. In addition, security is a very important aspect of software and I believe trying to find a solution to protect databases from inference attacks will give me valuable experience for my future career.

Ryan -

The reason I chose this project to be my 4th year project because it has the suitable knowledge and component of Software Engineering, for example the process of requirement gathering, user documentation, technical concept exploration, software architectural design, implementation, testing, and revision. Throughout this project, it

might also influence the decision of my future career study since the amount of database user had increasingly changed and most of the information are not securely stored.

Aleksandar -

The reason I chose the Confidentiality Preservation in Big Data Centres project is because it has significant components of Software Engineering within such as requirements, architectural and detailed design, implementation, verification and validation, and user documentation. On top of this, it also revolves around the topic of cyber security and could possibly lead into a career in the security sector.

Hasan -

I chose to be a part of the Confidentiality Preservation in Big Data Centres project because it has suitable components of Software Engineering within such as the development of requirements, architectural and detailed design, implementation, verification and validation, and user documentation. It piqued my interest as it could possibly lead into a career in the security sector which I believe is a great fit for my future in Software Engineering. The growing amount of big databases has increased the chance of privacy breaches, and tackling a solution to this problem is imperative and will provide me with great experience for the future.

Tashfiq -

The reason I found this to be suitable for my 4th year final project is because it utilizes great software engineering fundamentals that were taught/demonstrated throughout the undergraduate years. It utilizes the obtained knowledge of databases, queries, security/bugs, organized planning, team functionality, software requirements, software design, testing, and much more with strong programming skills. Another reason I chose this project is because I have a great interest in anything security based and I feel that a security protection against every possible inference attacks in big data centres is extremely interesting and difficult, which will be a challenge. This project has a great scope where it allows the protection for clients from attackers on personal confidential information. It would be an honor to complete this project, with excellence, with the strong team that I have plus the great instructor.

Contributions

Calvin

- Researched Cell Suppression method
- Added task breakdown

Ryan

- Researched database option
- Researched examples of inference attacks on different database

Aleksandar

- Researched benefits
- Added timetable
- Worked on background, objectives, methods

Hasan

- Added background
- Added objectives in-line with the project scope
- Lined up timetable with project team members schedules

Tashfiq

- Researched Polyinstantiation method
- Researched examples of inference attacks in background

Calvin Soong 

Ryan Zheng 

Aleksandar Savic 

Hasan Issa 

Tashfiq Akhand