Confidentiality Preservation in Big Data Centres

Supervisor: Jason Jaskolka
Team Size: Minimum 3, Maximum 4
Program of Study for Team Members:

CSE	SE	Comm	Biomed	EE
Yes	Yes	No	No	No

Description:
With more and more systems collecting enormous amounts of sensitive personal information, ensuring adequate protection of this information is of the utmost importance. When dealing with large and complex data centres, if proper care is not taken to design the data model and access control system to limit access to sensitive information to only authorized users, it may be possible for an attacker to conduct an *inference attack* to learn some information that they are not authorized to know. An inference attack is a process by which an attacker performs authorized queries and deduces (through reasoning) unauthorized information from the legitimate responses received. Such an attack results in a violation of the system's confidentiality policy.

This project aims to develop a solution to enforce the confidentiality policy in a big data centre. The goal is to find a way to automatically determine whether a system or user can access data fragments that may be linked together to form new knowledge that they are not authorized to know. For example, upon transactions to the data centre, the system may run a check to explore violations of the confidentiality policy. This project will also explore solutions for preventing these violations, such as raising warnings, redact pieces of information, or forbidding these transactions. The developed solution should not unduly interfere with the performance of the big data centre and the system operation.

This project will involve the development of the requirements, architectural and detailed design, implementation, verification and validation, as well as user documentation of a system that can support confidentiality preservation in big data centres. The solution should be configurable to enable users to specify particular confidentiality policies and mitigation solutions based on the particular use case. Possible applications of such a solution include e-Health (electronic health records), social networking (post history), and many more.

Objectives:

1. Identify a suitable application(s) and generate and obtain a suitable case study data set.
2. Develop a mechanism to specify the system confidentiality policy.
3. Implement mechanisms and software tools for identifying potential inference attacks violating the specified confidentiality policy.
4. Implement and deploy strategies to enforce the specified confidentiality policy.

Leveraged or Learned Knowledge:

- Software development
- Cybersecurity
- Reasoning

References:

References related to inference attacks and reasoning will be provided.

Prerequisites (in addition to those of SYSC-4907):
Strong programming skills, Strong interest in security and reasoning

Keywords:
Information Security, Confidentiality, Inference Attack, Reasoning, Big Data, Programming