Nuno Antunes, nmsa@dei.uc.pt
Nuno Cardoso, ncardoso@uc.pt

Assignment 1: Data Sharing with Encryption

October 14th @ 23:59

Objectives

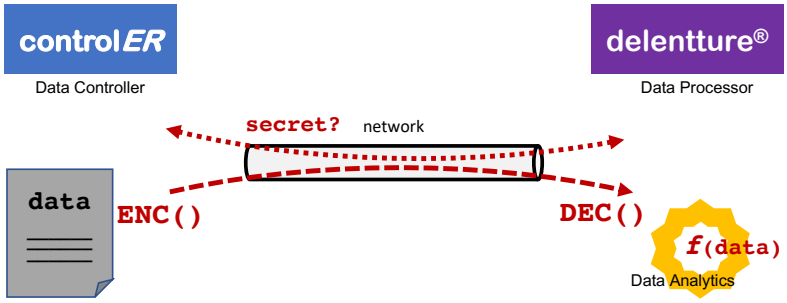
- Studying threats that affect data sharing scenarios.
- Understand how to use cryptography to protect the security of the data.
- Apply these techniques in the solution of concrete problems.

Preparation

Before starting the exercise, you should carefully read this assignment and understand what tasks are to be conducted and the goals to be achieved.
The assignment is prepared to be executed by **Groups of 2 elements**, in a total of 32 hours.

Problem statement

The company ControlER provides a service of consumer loans. Recently, the company noticed a significant raise in loan contract infringements, and it is worried about the current strategy in-place for assessing loan risk. ControlER does not have a data analysis/science team yet but obtaining insights on strategies to restrict loan approval is an urgent matter. As such, ControlER decided to contract the services of Delentture, a consulting company with highly specialized data science teams. The figure below depicts the scenario.



At this point, a major issue emerges: ***how will ControlER share the highly sensitive customer data with Delentture, for their data team to analyze and report on?***

In the exercise, you will design, implement, and evaluate a solution that will allow the ControlER organization to address this problem.

Resources

`ucstudent.uc.pt/2023_SP_Dataset.zip` includes two `.csv` files:

- `columns_description.csv` – describes the meaning of each of the columns of the dataset.
- `infringement_dataset.csv` – the content of the dataset to be used.

`ucstudent.uc.pt/2023_SP_Example.zip` includes an example of two parties communicating through a file, to support the development of the assignment.

Exercise: Encryption-based Solution

In this exercise you will implement the solution proposed by one of the ControlER engineers, which consists in sharing data in an encrypted CSV file.
To implement this solution, it will be necessary to create setup with two independent parties. Using the techniques learned in the course, the first (ControlER) sends the encrypted data through the network[1], while the second (Delentture) receives this data, decrypts, and checks its integrity and authenticity, and finally performs data analytics over this data.

For this, you will have to perform the following tasks:

1. Describe the threat model you are considering, i.e., who are the attackers you are considering and what are their capabilities.
2. Design the complete communication scheme including the selection of cryptographic algorithms and secret distribution schemes.
3. Implement the actual key exchange between the two parties of your system.
4. Implement the functions that use the selected algorithms and libraries for:
    a. encryption.
    b. communication.
    c. decryption.
    d. validation of integrity and authenticity.
5. Implement 2 data analysis to be performed by Delentture in the received data as follows:
    a. Each row in the dataset contains a loan application, with a flag (column 'infringed') stating whether the client infringed the contract at some point (=1), or did not infringe (=0). There is also information about past credit applications (columns starting with 'past_'. Briefly analyze the impact of these columns on the infringements.
    b. The second analysis should be conceived by you.
6. Execute the process and take note of the duration of the main activities.
7. Remove integrity and authenticity concerns from your communication scheme and repeat the process. Briefly discuss the differences.
8. Document and discuss all the obtained results in your report, including the results of your data analysis.

Deliverable

- A report containing:
    - the description of the threat model.
    - the complete design of your system and communication scheme.
    - the evaluation results, their analyses and discussion.
- The complete sources developed and necessary for the results obtained.

Questions

If you have questions about the scope of the work or any other aspect, please talk with the Professors of the course. You can do so face-to-face, by email, or using Skype.

[1] To avoid using 2 machines, this can be implemented using two independent processes running in localhost. You can use real sockets to communicate, or you can simulate this using binary files, as the example provided.