FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Assignment 1:
Data Sharing with Encryption

Segurança e Privacidade

Mestrado em Engenharia e Ciência de Dados

2022/2023

Duarte Emanuel Ramos Meneses – 2019216949 – duartemeneses@student.dei.uc.pt

Patrícia Beatriz Silva Costa – 2019213995 – patriciacosta@student.dei.uc.pt

Índice

Introdução

Nos dias que correm, o mundo é controlado por informação (dados). Como tal, para que não haja nenhuma desgraça na sociedade, é necessária a presença de segurança nas trocas de informação.

É com isso em vista que este trabalho prático procura simular a troca de informações confidenciais entre duas empresas. Para tal, são utilizados mecanismos de troca de chaves e de encriptação/desencriptação, de modo a garantir confidencialidade, integridade e autenticidade. Sem estes mecanismos, a comunicação entre as duas entidades estaria comprometida e exposta a possíveis atacantes.

Ao longo deste relatório vamos abordar o modelo de ameaça ao sistema e o esquema de comunicação. Vamos ainda analisar o impacte de alguns dados em empréstimos em incumprimento, bem como comparar o tempo de execução do nosso programa com e sem mecanismos de validação da autenticidade e da integridade.

1. Modelo de ameaça: quem são os invasores e quais são as suas capacidades

Atualmente, sendo a sociedade movida e controlada pela informação, os dados são dos recursos mais valiosos do mundo. Deste modo, a obtenção de informação é muito apetecível.

Com isto, a troca de dados entre a ControlER e a Delentture acarreta muitos riscos de segurança e privacidade. Um intruso pode colocar-se no meio da comunicação e intersetar ficheiros confidenciais.

Caso os ficheiros não estejam encriptados com um algoritmo que garanta autenticidade e integridade, o atacante pode alterar os dados sem que o suposto destinatário perceba. Já se não estiverem encriptados de todo, a tarefa de aceder aos dados por parte do invasor torna-se ainda mais fácil, já que a confidencialidade está comprometida.

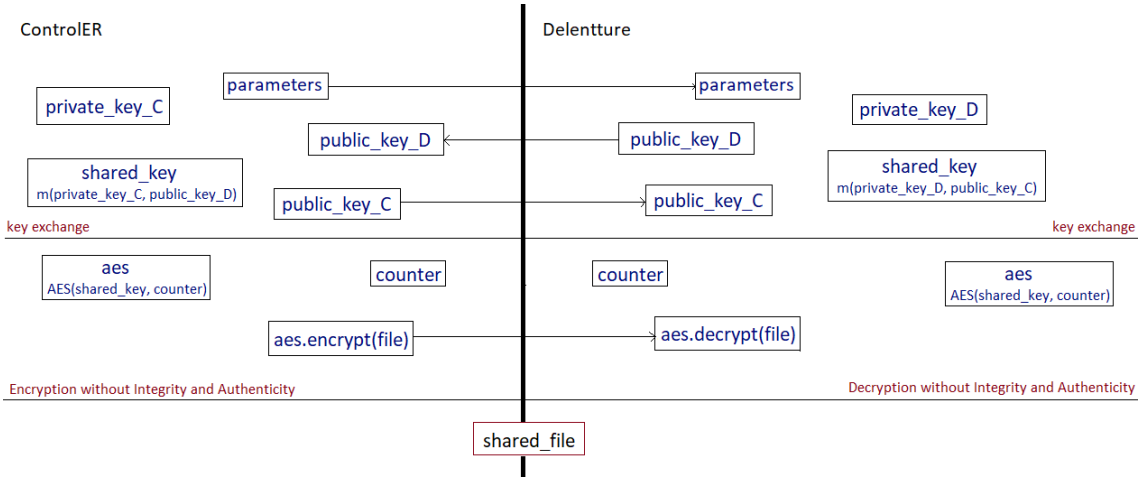
No entanto, encriptando os ficheiros, o invasor tem essa tarefa mais dificultada. Se, aliada à encriptação, o algoritmo garantir integridade e autenticidade, o invasor não consegue alterar os documentos sem que os verdadeiros interlocutores na comunicação reparem já que esses mecanismos de validação visam garantir que os ficheiros não são modificados por pessoas não autorizadas.

É importante referir que os atacantes podem ainda aceder ao sistema de ficheiros de uma máquina (ou mesmo à máquina em si), tendo acesso a documentos que não deviam. Neste caso, os ficheiros devem estar encriptados com os mecanismos devidos nas máquinas em si.

2. Desenho do esquema de comunicação

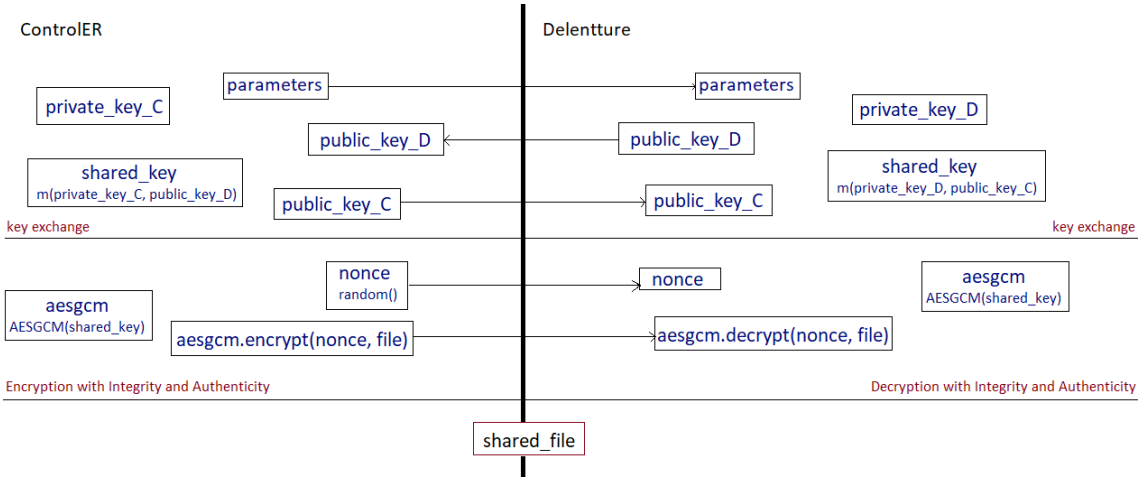
2.1. Sem validação da Autenticidade e Integridade

Decidimos utilizar o algoritmo de Diffie-Hellman para trocar a chave e o AES em modo CTR para encriptar o ficheiro.



2.2. Com validação da Autenticidade e Integridade

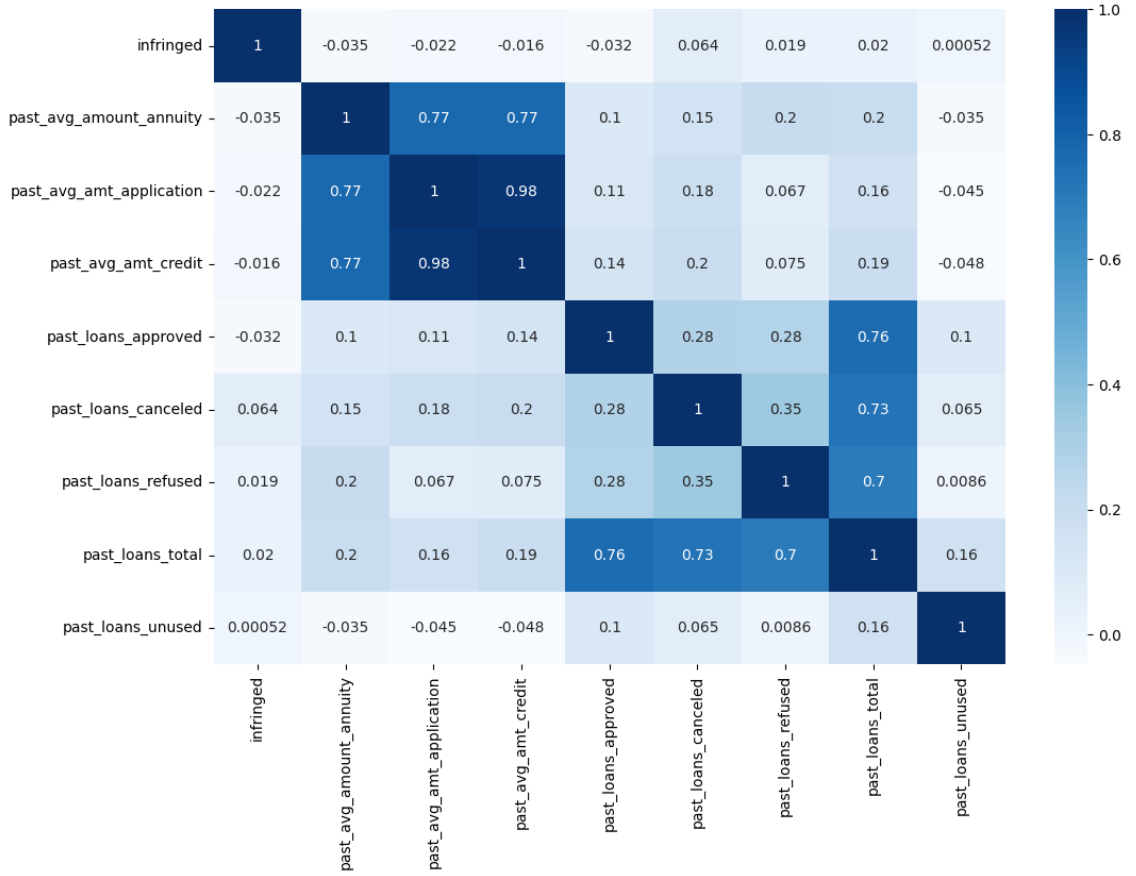
Já neste caso, optámos por utilizar o algoritmo de Diffie-Hellman para trocar a chave e o AESGCM para encriptar o ficheiro.



5. Análise de dados

5.1. Impacte das colunas "*past*" nos incumprimentos

Decidimos visualizar o impacte das colunas "*past*" nos empréstimos não cumpridos através da correlação entre essas colunas. O resultado é o seguinte:



Fica claro que a coluna "*past*" que mais impacte tem nos empréstimos em incumprimento é a coluna dos empréstimos passados cancelados. Podemos deduzir que estes foram cancelados precisamente por estarem em incumprimento ou por o cliente ter alguns nessa situação.

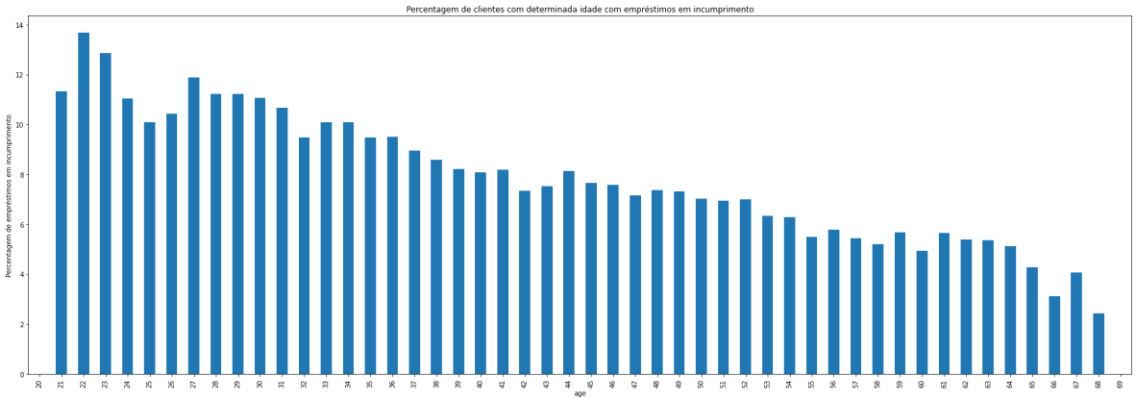
Embora com impacte inferior, as colunas relativas a empréstimos passados recusados e totais também têm alguma correlação com os em incumprimento. Podemos explicar os recusados uma vez que, talvez o tenham sido por, no passado, o cliente ter falhado em alguns empréstimos. Quanto ao número total, quantos mais empréstimos um cliente tem, mais provável é a existência de incumprimentos.

Por último, com ainda menos impacte, a coluna que diz respeito aos empréstimos aprovados mas não utilizados pelo cliente tem uma ligeira correlação com a coluna dos em incumprimento. Isto pode dever-se ao facto de o cliente não utilizar um empréstimo aprovado pois tem outros já em incumprimento.

Nenhuma outra coluna "*past"* tem impacte na coluna "*infringed"*.

5.2. Impacte da idade dos clientes nos empréstimos em incumprimento

Nesta secção, decidimos analisar o impacte da idade dos clientes nos empréstimos em incumprimento. Para tal, analisamos a percentagem de pessoas com determinada idade com empréstimos em incumprimento. Os resultados foram os seguintes:



Tal como era espectável, quanto mais novos são os clientes, a percentagem de empréstimos em incumprimento é superior. Podemos explicar isto, uma vez que com a idade se vai ganhando estabilidade financeira. Posto isto, quanto mais velho for um cliente, teoricamente, mais capacidade tem de pagar um empréstimo.

6. Tempo de execução

Para esta questão, decidimos realizar cinco medições de tempo para cada caso para evitar possíveis *outliers*. As medições podem ser consultadas em detalhe nos anexos "TimeWith.png" e "TimeWithout.png". Os resultados obtidos foram os seguintes (sendo Mx referente à medição número x):

Troca de chave + Encriptação sem validação da Autenticidade e Integridade					
Tempos	M1	M2	M3	M4	M5
	3,345	3,436	4,669	3,856	3,012
Média	3,6636				
Desvio Padrão	0,570217				

Troca de chave + Encriptação com validação da Autenticidade e Integridade					
Tempos	M1	M2	M3	M4	M5
	3,413	3,312	3,306	4,485	3,123
Média	3,5278				
Desvio Padrão	0,487655				

Analisando os resultados acima, percebemos que na medição dos tempos do algoritmo com validação da autenticidade e integridade existiram menos *outliers* que no outro caso. No entanto, isso não é relevante, uma vez que o desvio padrão, apesar de o comprovar, não tem um valor muito significativo.

Tendo medido cinco vezes para cada caso, consideramos que a média obtida espelha bem a duração de cada algoritmo.

À priori, esperávamos que um algoritmo com mecanismos de validação da integridade e autenticidade demorasse mais tempo visto ter de efetuar mais operações.

No entanto, não é isso que os resultados demonstram. Uma possível explicação para isto passa por a biblioteca utilizada não permitir paralelização no *counter* do AES-CTR, apesar de se esperar que acontecesse. Isto leva a que, ao contrário das expectativas, o algoritmo sem mecanismos de validação de autenticidade e integridade demore ligeiramente mais do que o que contém esse tipo de mecanismos.

Conclusão

Ao longo deste trabalho ficou claro que devemos sempre encriptar ficheiros com algoritmos que tenham mecanismos de validação da integridade e autenticidade. Isto leva a que haja uma maior segurança já que, mesmo que um invasor intersete o ficheiro, não consegue alterá-lo uma vez que não tem permissões para tal.

Percebemos também que, analisando o *dataset* disponibilizado, existem colunas referentes aos empréstimos passados que estão correlacionadas com os empréstimos em incumprimento. Olhando também para a idade de quem infringe, reparamos que existe uma maior percentagem de infratores nas gerações mais novas, o que pode significar uma menor estabilidade financeira nestas idades.

Em suma, este trabalho ajudou a colocar em prática a matéria lecionada nas aulas teóricas e práticas da cadeira, o que permitiu que consolidássemos e entendêssemos melhor os conceitos abordados em Segurança e Privacidade.

Referências

- Lutes, J. (2020, 2 de dezembro). *Correlation Is Simple With Seaborn And Pandas*. Medium. https://towardsdatascience.com/correlation-is-simple-with-seaborn-and-pandas-28c28e92701e - acedido em 8 de outubro 2022
- *Pandas - Data Correlations*. (s.d.). W3Schools Online Web Tutorials. https://www.w3schools.com/python/pandas/pandas_correlations.asp - acedido em 8 de outubro 2022
- *Pandas Correlation of Columns*. (s.d.). Spark by {Examples}. https://sparkbyexamples.com/pandas/pandas-correlation-of-columns/ - acedido em 8 de outubro 2022
- Gurav, S. (2022, 10 de agosto). *5 Pandas Group By Tricks You Should Know in Python*. Medium. https://towardsdatascience.com/5-pandas-group-by-tricks-you-should-know-in-python-f53246c92c94 - acedido em 8 de outubro 2022
- Lee, A. (2020, 10 de maio). *Making Plots with the Pandas groupby*. Medium. https://python.plainenglish.io/making-plots-with-the-pandas-groupby-ac492941af28 - acedido em 8 de outubro 2022
- *Plot the Size of each Group in a Groupby object in Pandas - GeeksforGeeks*. (s.d.). GeeksforGeeks. https://www.geeksforgeeks.org/plot-the-size-of-each-group-in-a-groupby-object-in-pandas/ - acedido em 8 de outubro 2022
- *Pandas GroupBy - Count occurrences in column - GeeksforGeeks*. (s.d.). GeeksforGeeks. https://www.geeksforgeeks.org/pandas-groupby-count-occurrences-in-column/ - acedido em 8 de outubro 2022
- *Run a basic correlation between two columns of a dataframe*. (s.d.). Stack Overflow. https://stackoverflow.com/questions/35095249/run-a-basic-correlation-between-two-columns-of-a-dataframe - acedido em 8 de outubro 2022
- *Welcome to pyca/cryptography — Cryptography 39.0.0.dev1 documentation*. (s.d.). Welcome to pyca/cryptography — Cryptography 39.0.0.dev1 documentation. https://cryptography.io/en/latest/ - acedido em 7 de outubro 2022
- Antunes N., (2022). Slides Teóricos, MECD 2022/23 - acedido em 14 de outubro 2022
- Cardoso N., (2022). Material Prático, MECD 2022/23 - acedido em 14 de outubro 2022