MPEI 2024/25
Relatório do Trabalho Prático

Ellen Sales - 117450 **Rodrigo Santos - 119198**

Prof. António Joaquim da Silva Teixeira

1 Introdução

1.1 Contextualização

Neste relatório, apresentamos o desenvolvimento de uma aplicação para análise de flows de dados em redes, com o objetivo de detectar comportamentos associados a ataques DDoS. Este projeto combina três componentes principais: um Classificador *Naïve Bayes*, um *Bloom Filter* e um *MinHash*, bem como um conjunto de testes que permite saber a precisão do código implementado.

1.2 Dataset

O dataset utilizado neste trabalho é baseado no **DDoS Dataset** (kaggle.com), com as devidas simplificações e filtragens que seriam pertinentes para o nosso caso de uso. Optámos por utilizar 10.000 linhas, obtidas aleatoriamente pelo código do ficheiro *dataset.py* e retirámos apenas as colunas de informação que considerámos relevantes.

1.3 Fluxo de Execução

Para a execução propriamente dita, utilizaremos a partição do dataset que reservámos para fazer testes. Assim, começamos por dar uso ao *Bloom Filter* para verificar se o Flow a analisar já foi identificado como maligno. Se não for, utilizaremos o Classificador *Naïve Bayes*, e *MinHash* para verificar semelhanças em relação a documentos já utilizados. O veredito será um ataque `ddos` apenas se ambos retornarem que é um ataque. Após dar o veredito, se o Flow for identificado como parte de um ataque DDoS, o seu identificador será adicionado ao *Bloom Filter*, de forma a torná-lo cada vez mais eficaz a cada iteração.

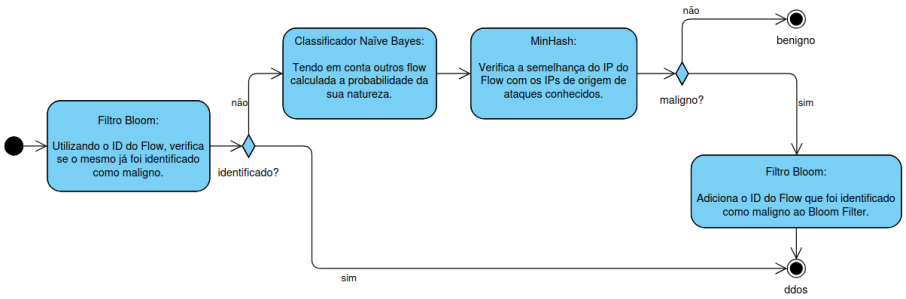

Figure 1: Fluxo de Identificação de IPs utilizando Bloom Filter, Naïve Bayes e MinHash.

2 Implementação

2.1 Bloom Filter

O **Bloom Filter** implementado segue as seguintes etapas:

1. **Inicialização:**
   - Vetor binário de tamanho 45000 inicializado.
   - 7 funções de hash geradas com um valor aleatório para cada uma.
2. **Inserção no Bloom Filter:**
   - IDs com classe `ddos` do treino são adicionados ao filtro usando as funções de hash.
3. **Verificação:**
   - Verifica se o ID do flow fornecido já foi analizado ou se sabemos a sua natureza. O seu valor será guardado na variável `status`, tendo valor `1` se pertencer ao Bloom Filter e `0` se não.
4. **Resultados:**
   - Se `status = 1`, classifica-se o flow como `ddos` (maligno). Se o ID não tiver sido adicionado ao Bloom Filter, o fluxo terá de ser submetido a análise para determinar a sua natureza.

2.2 Naïve Bayes

O classificador **Naïve Bayes** implementado utiliza o teorema de Bayes para classificar dados. Ele recebe o *conjunto de treino*, as *classes de treino* e o *conjunto de teste*.

Funcionamento:

1. **Identificação das classes:** As classes únicas do treino são identificadas e armazenadas.
2. **Cálculo das probabilidades a priori:** Determinadas pela proporção de exemplos de cada classe no treino.
3. **Cálculo das médias e variâncias:** Para cada classe, calcula-se a **média** e **variância** dos atributos, assumindo uma distribuição Gaussiana.
4. **Classificação:** Para cada amostra do teste:
   - Calculam-se as probabilidades condicionais $P(x|C)$ para cada classe.
   - Multiplicam-se pelas probabilidades a priori para obter $P(C|x)$.
   - A classe com maior probabilidade é atribuída como previsão.
5. **Resultado:**
   - **Previsões:** vetor com as classes atribuídas.
   - **Probabilidades:** matriz com as probabilidades de cada classe.

2.3 MinHash

O algoritmo **MinHash** foi implementado com as seguintes etapas:

1. **Inicialização:**
   - 100 funções de hash geradas com sementes aleatórias.
   - Número primo $2^{31} - 1$ utilizado para cálculo dos hashes.
2. **Geração de Shingles e Assinaturas:**
   - Shingles de tamanho 8 gerados para os IPs.
   - Assinaturas MinHash calculadas para IPs das classes `ddos` e `Benign`.
3. **Classificação:**
   - Para cada IP de teste:
     - Calcula-se a similaridade com as assinaturas das classes `ddos` e `Benign`.
     - A classe com maior similaridade média é atribuída.

4. **Resultados:**

    - Matriz com *shingles*, assinaturas, similaridades médias e classificação final.

Resultado:

- O algoritmo retorna a sua previsão (1 para *DDoS* e 0 para *Benign*) para cada IP analisado, baseada na comparação da similaridade de Jaccard calculada para os *shingles* da amostra em relação às duas categorias.

3 Testagem

Para os testes, o dataset com 10.000 linhas (5.000 *DDoS* e 5.000 *Benign*) foi dividido em treino (7.000 linhas) e teste (3.000 linhas), mantendo o equilíbrio entre as classes. A figura abaixo mostra os resultados de um dos testes feitos durante o desenvolvimento da aplicação.

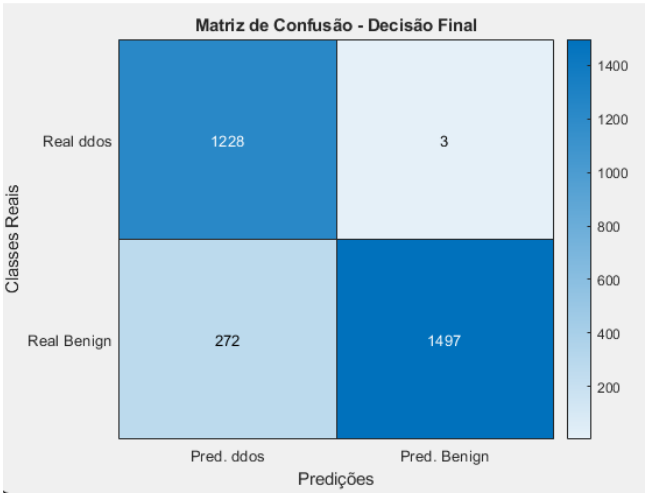

Figure 2: Matriz de Confusão - Decisão Final. A matriz apresenta a distribuição dos resultados entre as classes *Real ddos* e *Real Benign*, comparando com as previsões *Pred. ddos* e *Pred. Benign*. A precisão total é de aproximadamente **90,83%**.

4 Conclusão

Em geral, a aplicação desenvolvida apresentou resultados promissores na identificação de fluxos associados a ataques DDoS. A combinação dos três métodos – Bloom Filter, Naïve Bayes e MinHash – mostrou-se eficiente, em sua maioria com alta precisão (acima de 85%).