UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE – UERN
FACULDADE DE CIÊNCIAS EXATAS E NATURAIS – FANAT
DEPARTAMENTO DE INFORMÁTICA – DI

Davi Oliveira Rebouças

SIDAR - SISTEMA DE DETECÇÃO DE ATAQUES EM REDES

MOSSORÓ - RN

2021

SIDAR - Sistema de Detecção de Ataques em Redes

Relatório apresentado ao curso de Ciência da Computação da Universidade do Estado do Rio Grande no Norte como requisito da disciplina de Trabalho de Diplomação, sob a orientação do Prof. D.Sc Isaac de Lima Oliveira Filho.

MOSSORÓ - RN

2021

Davi Oliveira Rebouças

SIDAR - Sistema de Detecção de Ataques em Redes

Relatório apresentado como pré-requisito para obtenção do título de Bacharel em Ciência da Computação da Universidade do Estado do Rio Grande do Norte – UERN, submetida à aprovação da banca examinadora composta pelos seguintes membros:

Aprovado em: 13 / 05 / 2021

Banca Examinadora

Prof. D.Sc Isaac de Lima Oliveira Filho (Orientador)
Universidade do Estado do Rio Grande do Norte – UERN

Prof. D.Sc Rommel Wladimir de Lima
Universidade do Estado do Rio Grande do Norte – UERN

Prof. D.Sc Sebastião Emidio Alves Filho
Universidade do Estado do Rio Grande do Norte – UERN

SUMÁRIO

1 INTRODUÇÃO

Existem atualmente vários mecanismos de proteção de dados, dentre eles podemos citar os firewall (Stallings, 2013), que são mecanismos que bloqueiam determinadas portas (chamadas também de canais de serviços) de acesso externos, ou seja restringindo o acesso a determinados componentes de rede (máquinas, servidores, terminais, etc). Outro serviço importante na proteção da infraestrutura de rede, principalmente se esta permite tráfego de informações entre outros pontos, é o uso de VPN's (Virtual Private Networks)(Stallings, 2013). Nestas redes é possível criar um canal virtual de comunicação seguro entre hosts, utilizando o link de internet (desprotegido).

No entanto, apesar de serem métodos bem eficientes, ainda é possível que dentro da rede utilizada possam existir pessoas não autorizadas, por exemplo, hackers, crackers ou colaboradores despreparados ou mal orientados, que podem prejudicar a segurança dos dados e do funcionamento dos serviços de rede.

Considerando que não existem esquemas de proteção que sejam 100% seguros e resistentes a ataques, nos últimos anos a pesquisa por métodos que possam, além de detectar, prever quando um ataque poderá ocorrer, ou mesmo conseguir anteceder certos tipos de ataques, se tornou uma alternativa eficaz ao aumento de proteção dos dados de redes de computadores.

2 OBJETIVOS

Objetivo Geral:

Promover a predição e detecção de ataques em redes de computadores a partir de bases de dados do tipo .csv com informações gerais sobre tráfegos de redes, em um sistema baseado em algoritmos supervisionados de Inteligência Artificial.

3 METODOLOGIA

Inicialmente, foi necessário pesquisar sobre quais tipos de algoritmos de Inteligência Artificial eram os mais indicados para trabalhar com conjuntos de dados de tráfego em uma rede. Paralelo a isso, foi realizado um levantamento das possíveis bases de dados (abertas) que contenham informações (coletas) de redes com determinados tipos de ataques, por exemplo, ataques de negação de serviços, escâneres de redes, tentativas de acesso não autorizado, etc.

Neste sentido, após a escolha destas bases e dos algoritmos supervisionados escolhidos inicialmente, foram realizados experimentos sobre as bases escolhidas, de forma a configurar e aprimorar os parâmetros de bases, com objetivo de a aumentar a acurácia (taxa de acerto na predição) em função do tipo de base escolhida, e dos algoritmos selecionados. Neste passo, foi necessário reformular várias das categorias da base original .csv que continham valores não ideais para a linguagem de programação Python, ocorrendo então o processo de reclassificação dessas entradas de valores categorizados para valores numéricos (Figura 2).

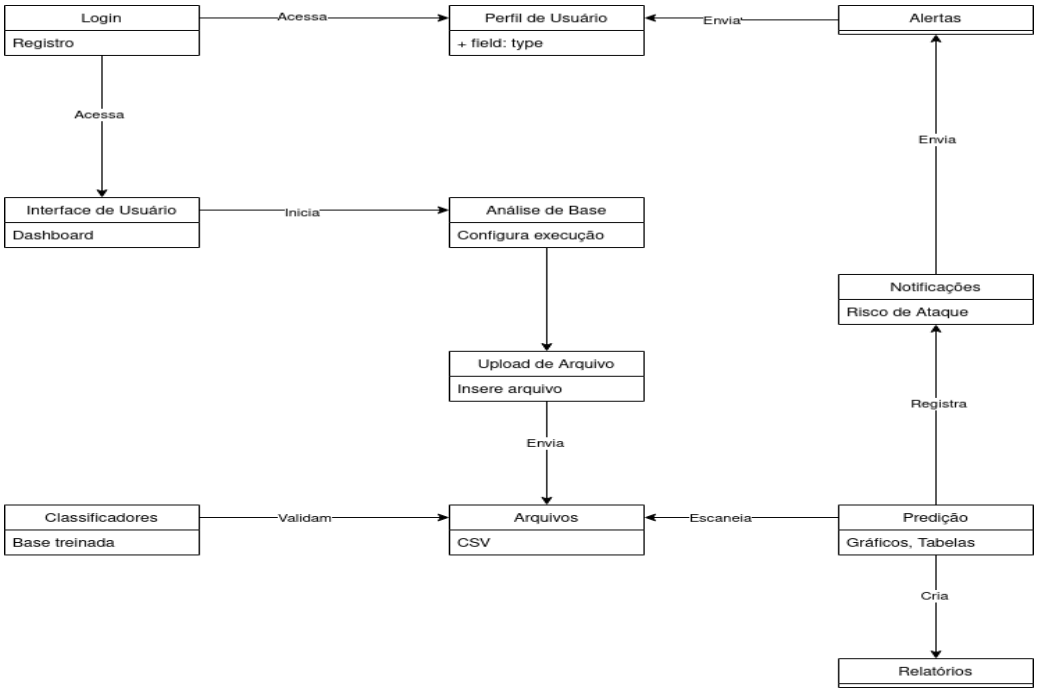
Em sequência, foi realizado o levantamento de requisitos do sistema proposto, em função das funcionalidades necessárias (Tabela 1) para que os usuários possam minimizar os riscos de ataques a seus serviços. Neste ponto, veio a parte de modelar (Figura 1), desenvolver e validar o sistema proposto, aplicando-o sobre outras bases de dados distintas.

Tabela 1 - Funcionalidades do Sistema

Cadastro e Login	Seção de autenticação dos usuários
Interface de Usuário	Dashboard com funções
Upload de Arquivos	Seção onde usuário envia arquivo para ser analisado (do tipo CSV)
Análise de Base	Base enviada é analisada pelos classificadores
Classificadores	Base treinada previamente com IA
Predição	Resultado da análise dos classificadores
Alertas	Gerados com base no resultado da predição

Fonte: Autoria Própria (2021)

Figura 1 - Diagrama UML de Sequência



Fonte: Autoria Própria (2021)

Figura 2 - Exemplo das Colunas da Base de Treino após Reclassificação

Src Port	Dst Port	Protocol	Flow Duration	Tot Fwd Pkts	Tot Bwd Pkts	TotLen Fwd Pkts	TotLen Bwd Pkts	Fwd Pkt Len Max	Fwd Pkt Len Min	Fwd Pkt Len Mean	Fwd Pkt Len Std	Bwd Pkt Len Max	Bwd Pkt Len Min	Bwd Pkt Len Mean	Bwd Pkt Len Std	Flow IAT Mean	Flow IAT Std
80	39470	6	4503818	4	4	935.0	356.0	935.0	0.0	233.7500	467.500000	356.0	0.0	89.000000	178.000000	6.434026e+05	1.668996e+06
52464	443	6	117375002	16	15	1151.0	7363.0	517.0	0.0	71.9375	136.408929	1430.0	0.0	490.866667	619.538523	3.912500e+06	1.479886e+07
80	56052	6	4444931	4	4	935.0	369.0	935.0	0.0	233.7500	467.500000	369.0	0.0	92.250000	184.500000	6.349901e+05	1.640111e+06
80	49846	6	4727750	4	4	935.0	351.0	935.0	0.0	233.7500	467.500000	351.0	0.0	87.750000	175.500000	6.753929e+05	1.731505e+06
80	60692	6	4343222	4	4	935.0	352.0	935.0	0.0	233.7500	467.500000	352.0	0.0	88.000000	176.000000	6.204603e+05	1.606079e+06

Fonte: Autoria Própria (2021)

4 DESCRIÇÃO DO SISTEMA

O SIDAR - Sistema de Detecção de Ataques em Redes - consiste em uma ferramenta on-line para auxiliar os usuários que trabalham em setores relacionados à segurança de rede. A ferramenta conta com a implementação de métodos de Inteligência Artificial para a predição de ataques de negação de serviço em redes de computadores.

Para utilizar este sistema, a base coletada pelo usuário deve estar formatada de acordo com as orientações no próprio sistema. A Inteligência Artificial, aplicada à ferramenta, analisa a probabilidade de ataque de negação de serviço que sua rede pode ter sofrido em função da base coletada pelo usuário.

O sistema conta com cadastro e autenticação via login, análise prévia da base, funções de configuração e estatísticas da base de dados. O sistema permite que o usuário realize o *upload* da base de dados em formato .csv e a partir da função de análise, receba um feedback de acordo com o índice de comprometimento da base em relação ao ataque de negação de serviço.

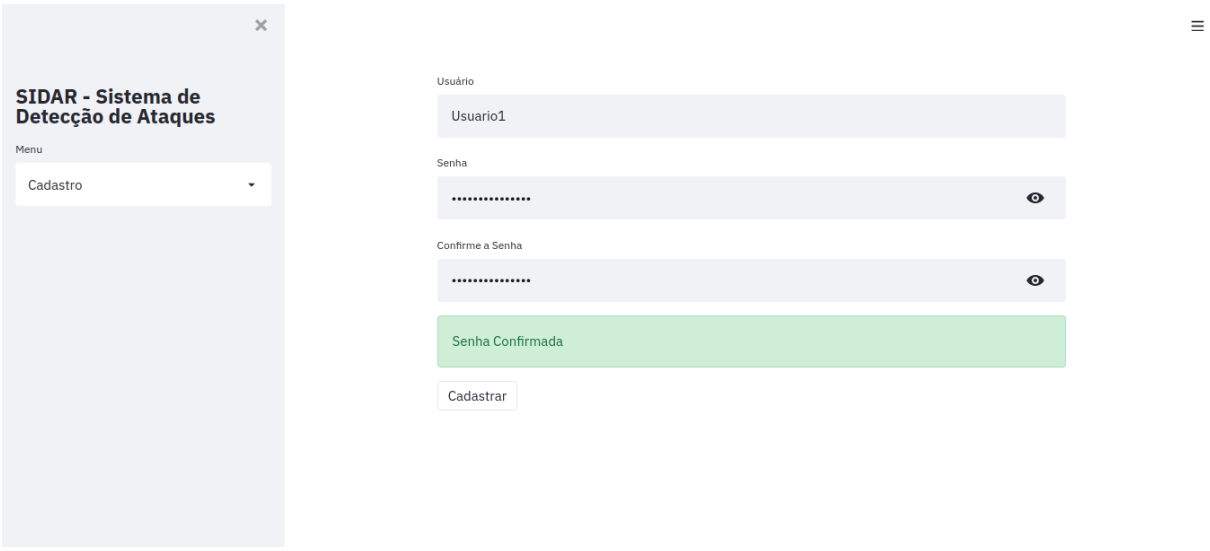
O sistema SIDAR foi criado na linguagem de programação Python e com a utilização dos frameworks Django e Streamlit, além de ferramentas de testes como Jupyter Notebook e Google Colab.

Na Figura 3, vê-se a tela inicial do sistema, onde o usuário pode ler um pequeno resumo do que se trata a ferramenta e ser direcionado ao menu lateral, onde pode se cadastrar (Figura 4) ou fazer login.

Figura 3 - Tela Inicial do Sistema

Fonte: Autoria Própria (2021)

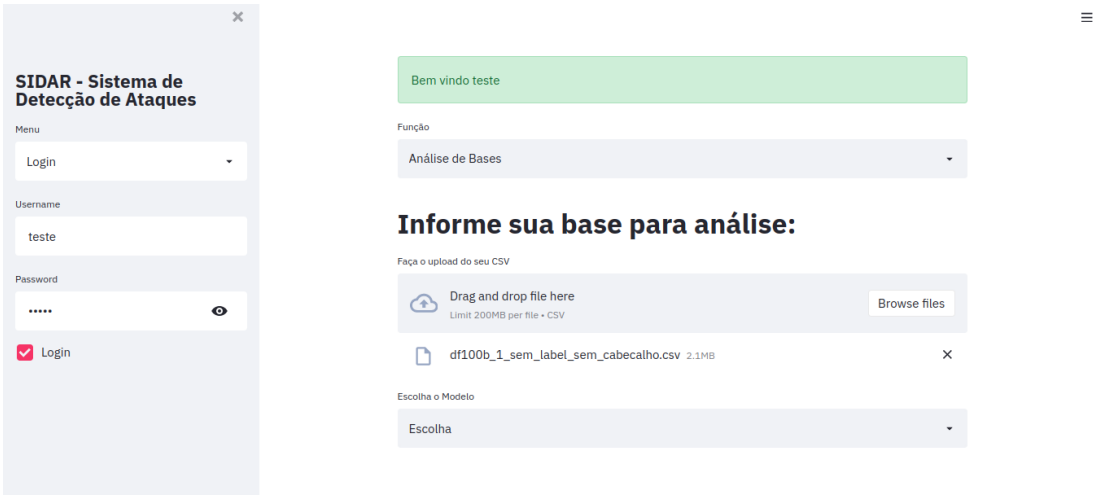
Figura 4 - Tela de Cadastro do Usuário



Fonte: Autoria Própria (2021)

Nas Figuras 5 e 6 vê-se as telas do processo de Análise de Bases, onde o usuário faz *upload* do seu arquivo, que é posteriormente analisado pelos classificadores e assim é dado uma determinada porcentagem com a possibilidade de ataque neste arquivo. Junto a esta porcentagem também é determinado através de um relatório se aquela porcentagem corresponde a um padrão propício a ataques ou não. As porcentagens de alerta foram demarcadas deliberadamente e seguem um padrão de 3 níveis com a maioria ou minoria da base com probabilidade de ataque para definir o nível de segurança daquele resultado.

Figura 5 - Tela de Análise de Bases



Fonte: Autoria Própria (2021)

Figura 6 - Tela de Resultado de Análise

Fonte: Autoria Própria (2021)

5 RESULTADOS

Foram feitos estudos a respeito dos algoritmos de inteligência artificial e um levantamento sobre os tipos de bases de dados relacionadas a ataques de redes e serviços, fomentando assim um aprendizado sobre pré-processamento e adaptação de bases, aprimorando assim o conhecimento em configuração, aplicação e treinamento de algoritmos de predição sobre bases de dados.

Ao final obteve-se a ferramenta SIDAR, que produziu resultados satisfatórios na prevenção e predição da possibilidade de ataques de negação de serviço em todas as bases de testes selecionadas.

Estão disponíveis três algoritmos de I.A para o usuário, que são:

- Algoritmo da Árvore de Decisão (Decision Tree): as árvores de decisão classificam instâncias partindo da raiz da árvore para algum nó folha que fornece a classe da instância.
- Algoritmo do Vizinho mais Próximo (K-Nearest Neighbour): o vizinho mais próximo é um método matemático o qual calcula a distância de dois elementos de acordo com sua diferença euclidiana.
- Algoritmo de Rede Neural (Neural Network): no algoritmo de rede neural os nós interconectados funcionam como os neurônios do cérebro humano. Usando camadas, elas podem reconhecer padrões escondidos e correlações em dados brutos.

Espera-se dar continuidade ao projeto e entre os próximos passos está a criação de uma API para o uso automatizado e agendado do sistema. Além disso, espera-se submeter os resultados a simpósios, congressos e revistas na área de segurança de redes.

REFERÊNCIAS

IDS 2018, A Realistic Cyber Defense Dataset (CSE-CIC-IDS2018). Amazon Web Services. Disponível em: https://registry.opendata.aws/cse-cic-ids2018/. Acesso em: Fevereiro de 2021.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. The MIT Press. 2016.Filho, Hayrton Rodrigues Prado. Os riscos e os prejuízos causados pela vulnerabilidade da segurança da informação. Disponível em: https://revistaadnormas.com.br/2018/05/29/os-riscos-e-os-prejuizos-causados-pela-vulnerabilidade-da-seguranca-da-informacao/. Acesso em: Abril de 2021

W. Stallings, Cryptography and Network Security: Principles and Practice (6th Edition), Prentice Hall, 2013.

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA ECONOMIA
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512021000869-3**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1° de janeiro subsequente à data de 29/04/2021, em conformidade com o §2°, art. 2° da Lei 9.609, de 19 de Fevereiro de 1998.

Título: SIDAR - Sistema de Detecção de Ataques em Redes

Data de publicação: 29/04/2021

Data de criação: 28/04/2021

Titular(es): UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE - UERN

Autor(es): DAVI OLIVEIRA REBOUÇAS; ISAAC DE LIMA OLIVEIRA FILHO; JOÃO ROBERTO DE ARAÚJO MENDES; EMÍDIO LOPES DE SOUZA NETO

Linguagem: PYTHON

Campo de aplicação: IF-02; IF-07; IF-10

Tipo de programa: GI-01; IA-01; PD-01

Algoritmo hash: SHA-512

Resumo digital hash:
f045a67b1867220eebd1784b67bc5ebd139aa1fa8fde9cb8718c049356f6b4bb70fa337f38d6ac2d9130d23463f2a57c9c6
64fc3c81878014e380457194174c0

Expedido em: 11/05/2021

Aprovado por:
Carlos Alexandre Fernandes Silva
Chefe da DIPTO