Tamara de Castro Lacerda

Análise de características que possibilitam a identificação de ataques DNS Spoofing através de técnicas de Reconhecimento de Padrões em uma LAN com sobrecargas de rede

Diamantina - MG

14 de setembro de 2017

Tamara de Castro Lacerda

Análise de características que possibilitam a identificação de ataques DNS Spoofing através de técnicas de Reconhecimento de Padrões em uma LAN com sobrecargas de rede

Trabalho de Conclusão de Curso apresentado ao Curso de Sistemas de Informação da Universidade Federal dos Vales do Jequitinhonha e Mucuri - UFVJM, como parte dos requisitos exigidos para a obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal dos Vales do Jequitinhonha e Mucuri – UFVJM

Faculdade de Ciências Exatas

Departamento de Computação

Orientador: Prof. MSc. Eduardo Pelli

Diamantina - MG

14 de setembro de 2017

Monografia de projeto final de graduação sob o título "Análise de características que possibilitam a identificação de ataques DNS *Spoofing* através de técnicas de reconhecimento de padrões em uma LAN com sobrecargas de rede", defendida por Tamara de Castro Lacerda e aprovada em 14 de setembro de 2017, em Diamantina, Minas Gerais.

Banca Examinadora:

Prof . MSc.Eduardo Pelli
Orientador

Prof. Dr.Alessandro Vivas Andrade

Prof. MSc.Rafael Santin

Henrique Carvalho Fonte Boa

Agradecimentos

A minha vó Alzira, por ser minha maior fonte de inspiração, por todo amor e dedicação.

À toda minha família, em especial aos meus pais, Érika e José Otaviano (Jacaré), pelo apoio e suporte, tão importantes nessa conquista.

Ao meu orientador Eduardo Pelli, pela oportunidade, confiança e suporte.

Ao Henrique, por todo incentivo e conhecimento compartilhado.

A todos meus amigos, em especial as moradoras da república, Jeanne, Maria Eduarda e Nayara, que me acolheram e foram como uma família em Diamantina para mim, ao Dudu, por ser meu constante motivo de risadas e amizade e, sem esquecer dos que trilharam essa jornada comigo e que hoje, mesmo longe, não deixam de ser importantes e exemplos para mim: Carlos Eduardo (Cadu) e Marcus Vinícius (Chiclete).

Ao Filipe pelo companheirismo e paciência, dividindo comigo todos os momentos nessa reta final.

Agradeço aos técnicos do Departamento da Computação, juntamente aos professores, ao DTI Desenvolvimento de Sistemas, a NextStep Empresa Jr. de Sistemas de Informação e ao INSS Gerência Executiva, responsáveis pelos maiores aprendizados e imprescindíveis para a minha graduação.

"Do... or do not. There is no try."
(Mestre Yoda)

Resumo

LACERDA, T. C. **Análise de características que possibilitam a identificação de ataques DNS Spoofing através de técnicas de Reconhecimento de Padrões em uma LAN com sobrecargas de rede**. Diamantina, 2017. 46. Trabalho de Conclusão de Curso Superior em Sistemas de Informação. Faculdade de Ciências Exatas, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2017.

O *Domain Name System* (DNS) é uma parte essencial da infraestrutura da Internet, alvo de constantes ataques cibernéticos. O ataque DNS *Spoofing* explora uma vulnerabilidade do cache do DNS, que armazena a primeira resposta à solicitação de tradução de domínio em IP como verdadeira. Em uma rede local, o computador do atacante consegue emitir uma resposta falsa à requisição rapidamente, envenenando o cache, que passa a reconhecer o IP do atacante como confiável. A aplicação de técnicas de reconhecimento de padrões, através das técnicas de seleção de características *F-Score*, Coeficiente de correlação de *Pearson* e o classificador SVM (*Support Vector Machine*), possibilita a identificação de ataque DNS *Spoofing* com alta precisão e acurácia. O presente trabalho teve como objetivo analisar se as características que possibilitam identificar um ataque DNS *Spoofing* através da aplicação das técnicas de reconhecimento de padrões, são afetadas pela sobrecarga da rede local. Constatou-se no decorrer do trabalho, que as métricas de avaliação supra são complementares e possibilitam a análise consistente dos dados. Verificou-se então que as sobrecargas simuladas na rede local não afetam de forma significativa as características relevantes na identificação de um ataque *Spoofing*.

Palavras-chave: DNS *Spoofing*; reconhecimento de padrões; redes de computadores.

Abstract

LACERDA, T. C. **Characteristics Analysis that makes possible identifying DNS Spoofing attacks through pattern Recognition Techniques in a LAN with network overcharge**. Diamantina, 2017. 46. Trabalho de Conclusão de Curso Superior em Sistemas de Informação. Faculdade de Ciências Exatas, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2017.

The Domain Name System (DNS), is an essential part of Internet infrastructure, constantly being the target of cybernetic attacks. The DNS Spoofing attack explores a vulnerability in DNS cache, that stores the first answer to the request of domain translation in IP as true. In a Local Area Network, the attacking computer can issue a fast false answer to the request, poisoning the cache, which begins to recognize the attacker's IP as trustworthy. The application of pattern recognition techniques, through the selection of F-Score characteristics, Pearson coefficient of correlation and SVM (Support Vector Machine) classifier, enables the identification of DNS Spoofing attacks with high precision and accuracy. This final paper has the objective to analyze if the characteristics that allow to identify a DNS Spoofing attack through the application of the pattern recognition techniques are affected by the overcharge of the local area network. At the end of the experiments, it was found that the methods of pattern recognition techniques above are complementary, and enables a consistent data analysis. With the exception of the speed average, it was verified that the simulated overcharges in local area network does not affect significantly in the relevant characteristics in identifying a DNS Spoofing attack.

Keywords: DNS *Spoofing*;pattern recognition;computer networks.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina
DNS	Domain Name System
IP	Internet Protocol
LAN	Local Area Network - Rede Local
PC	Personal Computer - Computador Pessoal
RNA	Rede Neural Artificial
RTT	Round Trip Time
SO	Sistema Operacional
SVM	Support Vector Machines - Máquinas de Vetores de Suporte
TTL	Time to Live
TCP/IP	Transmission Control Protocol/Internet Protocol
DoS	Denial of Service

Sumário

1 Introdução

A velocidade com que a conectividade digital está mudando a vida das pessoas em todo o mundo é cada vez mais visível. Segundo WEARESOCIAL (2017), o número de usuários que utilizam a Internet é de 3.77 bilhões, mais da metade da população global, com índice de 10% de aumento sobre o último ano. Os usuários que utilizam comercio eletrônico totalizam 1.61 bilhões de pessoas, e os adeptos a redes sociais 2.80 bilhões.

O crescente número de pessoas que utilizam a Internet implica em mais dispositivos conectados compartilhando informações. Para que os usuários da Internet acessem a web, eles devem ser capazes de se identificar. Existem duas maneiras de identificação dos hosts[1]: pelo nome de domínio e pelo endereço IP. As pessoas preferem utilizar o nome de domínio, enquanto os roteadores preferem os endereços de tamanhos fixos de IP. Para conciliar tais preferências, foi criado um serviço de diretório que traduz nomes de domínios para os endereços de IP. Esta é a principal tarefa do sistema de nomes de domínio na Internet (DNS) (KUROSE; ROSS, 2010).

O número de crimes que exploram as vulnerabilidades de rede, aumenta de forma proporcional ao número de usuários digitais. Quanto mais pessoas disponibilizam suas informações na web, maior o número de usuários maliciosos atrás destas informações. Segundo SYMANTECCORPORATION (2016), 684.4 milhões de pessoas foram afetas por cibercrimes[2] no mundo, gerando um prejuízo financeiro de US$125.9 bilhões.

O DNS é uma parte essencial da infraestrutura da Internet, e por isso é alvo de constantes ataques visando a negação do serviço, ou o redirecionamento das informações. O ataque DNS *Spoofing* consiste na interceptação do tráfego entre o cliente e o roteador *gateway*, contaminando o cache do servidor com resposta falsa a uma solicitação, redirecionando a resposta a um IP falso em benefício do atacante (MOHAMMED et al., 2016). A identificação desse tipo de ataque não é facilmente perceptível, acarretando em grandes prejuízos as vítimas.

O estudo de reconhecimento de padrões pode ser aplicado em duas categorias: estudo de todos os organismos vivos, observando o modo de desenvolvimento e aprimoramento de suas capacidades em reconhecer padrões e a aplicação e desenvolvimento de teorias e técnicas na criação de máquinas capazes de reconhecer padrões semelhante aos seres vivos (TOU; GONZALES, 1974).

Segundo Carvalho e Pelli (2017) a aplicação de técnicas de reconhecimento de padrões com a aplicação de técnicas de seleção de características como *F-Score*, Coeficiente

[1] Computador ligado a uma rede.
[2] Delitos que vão de atividades criminosas contra dados até infração de conteúdo copyright.

de Correlação de *Pearson* e a técnica de classificação de característica SVM é possível predizer com 98,33% de acurácia novos ataques de DNS *Spoofing*.

Considerando que a variação no número de requisições de serviços aumenta a carga da rede, foram propostos diferentes experimentos para analisar se as características identificadas no ataque DNS *Spoofing* são afetadas pelas sobrecargas da rede interna.

1.1 Objetivo Geral

Este trabalho teve como objetivo analisar através da aplicação de técnicas de reconhecimento de padrões, se as características que possibilitam a identificação de ataques DNS *Spoofing* são afetadas pela sobrecarga da rede local.

1.1.1 Objetivos Específicos

- Construção de um *testbed*[3] para a reprodução dos experimentos de sobrecarga da rede local e simulação de ataque DNS *Spoofing*;
- Aprender sobre scripts *Bash*, manipulação do R em conjunto com o RStudio para aplicar as técnicas de reconhecimento de padrões, uso do SO Kali Linux e módulo *Ettercap*;
- Estudo dos conceitos e teorias envolvidos neste projeto, como a aplicações das técnicas de seleção *F-Score* e Coeficiente de correlação de Pearson e classificação de características através do classificador SVM;
- Avaliação dos resultados após aplicar as técnicas de reconhecimento de padrões em diferentes experimentos de sobrecarga da rede.

[3] Ambiente controlado de testes

2 Revisão de Literatura

2.1 Redes de Computadores

Uma rede é um conjunto de dispositivos (normalmente conhecido como nós), conectados por links[1] de comunicação. Um nó pode ser um computador, uma impressora ou um dispositivo de envio e/ou recepção de dados, que estejam conectados a outros nós da rede. Uma rede deve ser capaz de atender a um certo número de critérios, sendo os mais importantes (FOROUZAN, 2009):

- Desempenho: em geral é preciso de mais capacidade de vazão e menos atraso;
- Confiabilidade: é medida pela frequência de falhas, pelo tempo que um link leva para se recuperar de uma falha e pela robustez da rede;
- Segurança: proteção ao acesso não autorizado de dados, proteção dos dados contra danos e a implementação de políticas e procedimentos para recuperação de violações e perdas de dados.

As redes locais, muitas vezes chamadas de LAN's, são redes privadas contidas em um único edifício ou campus universitário com até alguns quilômetros de extensão. São amplamente usadas para conectar computadores pessoais e estações de trabalhos em escritórios e instalações industriais de empresas, permitindo o compartilhamento de recursos e a troca de informações. As LAN's têm características que as distinguem de outros tipos de rede (TANENBAUM, 2003):

- Tamanho: As LAN's têm um tamanho restrito, o que significa que o pior tempo de transmissão é limitado e conhecido com antecedência;
- Tecnologia de transmissão: A tecnologia de transmissão das LAN's quase sempre consiste em um cabo, ao qual todas as máquinas estão conectadas, como acontece com linhas telefônicas compartilhadas que eram usadas em áreas rurais. As LAN's têm baixo retardo e cometem pouquíssimos erros;

2.1.1 Internet

A Internet é uma rede de computadores que interconecta milhares de dispositivos computacionais ao redor do mundo (KUROSE; ROSS, 2010).

Na década de 1970, houve uma revolução na rede de computadores: o conceito de Internet. Muitos pesquisadores estudaram a comutação de pacotes em busca de uma tecnologia simples de comutação de pacotes que pudesse atender a todas as necessidades.

[1] Um link é um caminho de comunicação que transfere dados de um dispositivo a outro.

Em 1973, Vintor Cerf e Robert Kahn propuseram o desenvolvimento de um conjunto de padrões para tal interconexão, e o resultado ficou conhecido como família TCP/IP de protocolos da Internet. A habilidade do TCP/IP de tolerar novas redes de comutação de pacotes é a maior razão para a evolução contínua das tecnologias de comutação de pacotes. À medida em que a Internet cresce, computadores se tornam mais poderosos e aplicações enviam mais dados (COMER, 2001).

A Figura 1 apresenta as 5 camadas da arquitetura TCP/IP, sobre a qual a Internet funciona.

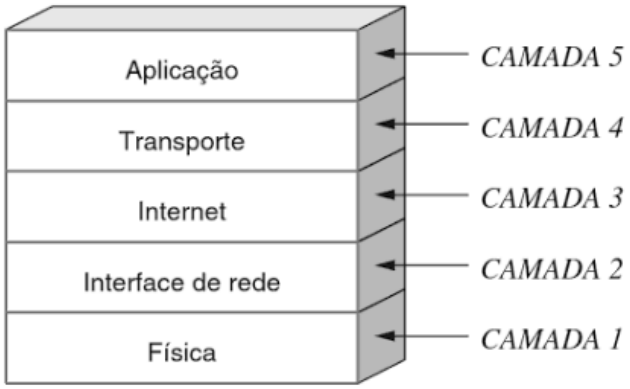

Figura 1 – Pilha de protocolos TCP/IP
Fonte: (COMER, 2001)

Dentro de uma rede TCP/IP, cada host recebe um endereço IP único que o identifica na rede. Cada endereço IP é dividido em duas partes: um prefixo e um sufixo. Um prefixo identifica a rede física à qual o host está conectado e um sufixo IP identifica um computador específico na rede. É importante que o esquema de endereços IP garanta duas propriedades (COMER, 2001):

- A cada computador é atribuído um endereço único (isto é, um endereço único nunca é atribuído a mais de um computador);
- Embora as atribuições do número de rede devam ser coordenadas globalmente, os sufixos podem ser atribuídos localmente sem uma coordenação global.

2.1.2 DNS

Desde o início da Internet, soluções foram utilizadas para evitar que pessoas tivessem que decorar o endereço da máquina para qual a comunicação tivesse de ser feita. Uma solução inicial foi criar uma tabela que contivesse um mapeamento entre o nome de um computador e o seu endereço. Assim, uma pessoa teria que apenas que decorar que o nome do servidor de e-mail da universidade era servidoremail, e não 200.156.84.209, por exemplo. Essa tabela, presente em cada computador da rede, era localizada em arquivos chamados de HOST.TXT. O crescimento do número de computadores ligados à Internet,

aumentou bastante o tamanho dessa tabela, o que impossibilitou sua administração de forma satisfatória. A alternativa adotada foi então criar um sistema de tradução de nomes em endereços (e vice-versa), que fosse hierárquico e distribuído. Esse sistema de tradução é conhecido como DNS (COSTA, 2007).

Um modelo de consulta ao DNS é apresentado na Figura 2.

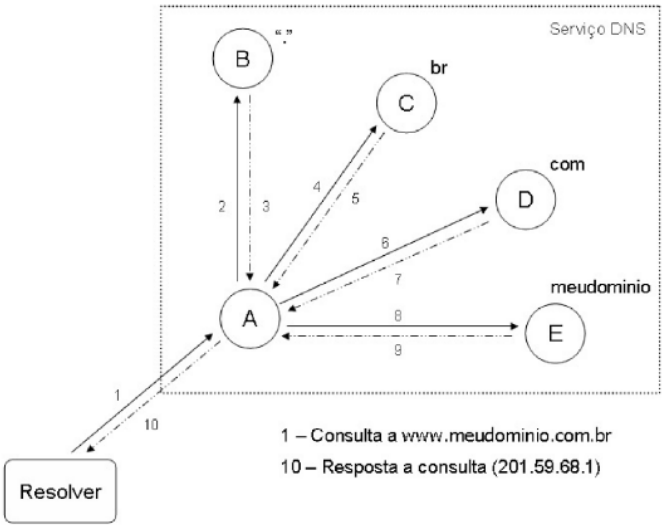

Figura 2 – Consulta típica de DNS
Fonte: (COSTA, 2007)

O cliente (resolver) pergunta ao servidor de DNS configurado em seu protocolo TCP/IP, qual o endereço de www.meudominio.com.br. O servidor recursivo de DNS recebe a consulta e aciona outros servidores para auxiliá-lo na pesquisa de forma iterativa. Quando o servidor autoritário (tem autoridade sobre o domínio).com, recebe a pesquisa este responde com endereço IP do host www.meudominio.com.br. O servidor DNS local retorna para o cliente o endereço IP, e o usuário consegue abrir uma conexão e visualizar a página.

Para melhorar a eficiência das consultas DNS, utilizam-se mecanismos de armazenamento temporário chamado *cache*. Com isso, consultas previamente realizadas com sucesso, de um domínio completo ou parte de um domínio, não precisam ser refeitas, uma vez que essas consultas ficam armazenadas em uma base de dados local (COSTA, 2007).

2.1.3 Vulnerabilidades do DNS

O DNS está sujeito a vários ataques, dentre os mais conhecidos (AL.HAJERI, 2000-2002):

- Transferência de zona: Uma zona é uma fonte autoritária de informações sobre cada nome de domínio DNS incluído na zona. Quando um servidor DNS é adicionado à rede e é configurado como um novo servidor secundário para uma zona existente,

ele executa uma transferência completa inicial da zona para obter e replicar uma cópia completa de registros de recursos da zona. Ao executar uma solicitação de transferência de zona, o invasor pode revelar o registro completo do DNS, incluindo endereços IP, locais de domínio, serviços de domínio, que podem ajudar o invasor a mapear a estrutura da rede.

- Negação de serviço (DoS): Em um ataque como SYN *flood*, o atacante gera a negação de serviço enviando um grande fluxo ininterrupto de pacotes de solicitação de DNS para a porta de serviço 53 do DNS de destino.
- *Buffer overflow*: Os atacantes podem causar um estouro de *buffer* no DNS ao emitir comandos contendo argumentos inesperados ou excessivamente longos. Isto ocorre devido a uma fraca codificação de software que permite que os invasores insiram código executável na memória. O estouro de *buffer* permitiria que um invasor executasse comandos no servidor do DNS a nível de privilégio.
- Envenenamento de *cache*: É um ataque de nível informacional, que responde ao servidor de domínio, com mais do que a resposta correta. Para melhorar o desempenho, os servidores DNS tentam "armazenar em *cache*"nomes de consultas localmente no sistema. Eles verificam todos os pacotes que entram no sistema em busca de uma resposta (cada pacote contém uma seção de consulta e resposta). Os servidores então lembram-se dessas respostas por um curto período de tempo no caso de qualquer outra pessoa precisar dessa informação. O problema óbvio é que alguém pode mentir. Em particular, alguém poderia enviar uma consulta ao servidor DNS que contenha informações adicionais, ou respostas falsas. Os servidores mais antigos aceitam essa informação, e armazenam em cache, fornecendo como resposta a qualquer outra pessoa que perguntar (novos servidores de DNS corrigiram isso).

2.1.4 DNS Spoofing

O DNS *Spoofing* é um termo utilizado quando um servidor DNS aceita e usa informações incorretas de um host que não possui autoridade para fornecer essa informação. A falsificação de DNS é de fato um envenenamento de cache malicioso onde dados forjados são inseridos no cache dos servidores de resolução de nomes. Os ataques de falsificação podem causar sérios problemas de segurança para servidores DNS vulneráveis, por exemplo, fazendo com que os usuários sejam direcionados para sites errados na Internet (AL.HAJERI, 2000-2002).

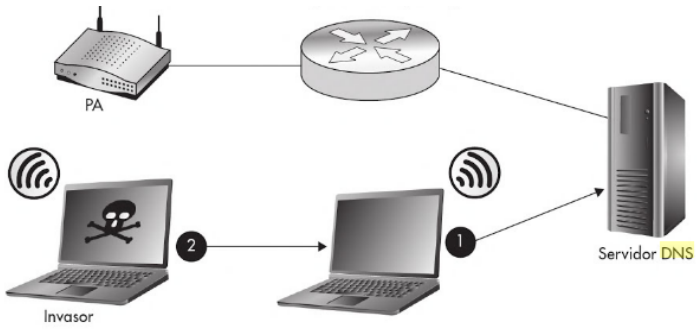

Figura 3 – Ataque DNS *Spoofing*
Fonte:(WRIGHTSON, 2014)

Na Figura 3: 1) Representa o cliente solicitando um IP de um domínio ao servidor DNS; 2) Antes de o servidor DNS conseguir responder, o atacante envia uma resposta alegando ser o servidor de DNS.

Para realizar um ataque de DNS *Spoofing*, o invasor deve estar conectado em qualquer lugar entre a vítima e o servidor DNS, e ele deve ser capaz de analisar todo o tráfego da vítima. Se a vítima e o atacante estiverem conectados a um hub, o invasor pode facilmente monitorar todo tráfego usando um dos vários *sniffers* disponíveis na Internet. O atacante terá acesso a todo tráfego do DNS, e será capaz de sequestrar a sessão do DNS e enviar uma resposta falsa sobre o IP real do endereço que a vítima está tentando acessar. Isso fará com que a vítima seja redirecionada para o destino que o atacante deseja (AL.HAJERI, 2000-2002).

2.2 Reconhecimento de Padrões

Reconhecimento de padrões pode ser definido como a categorização de dados de entrada dentro de classes identificáveis via extração de características significantes ou atributos com detalhes relevantes. O objetivo fundamental de um sistema de reconhecimento de padrões pode ser a classificação, como também a regressão (DUDA; HART; STORK, 2001).

Um sistema de reconhecimento de padrões pode ser dividido em algumas etapas: extração de características, pré-processamento, seleção e classificação.

Características são quaisquer medidas extraíveis de um padrão que podem contribuir para a classificação, sendo que as mesmas podem ser representadas por valores contínuos, que são valores mapeados de toda a população, ou discretos, valores mapeados de amostras da população (BARANOSKI, 2005).

O processo de extração de características consiste em mapear um problema específico em outro espaço de menor dimensão. Dessa forma, visa-se encontrar variáveis mais

representativas, levando em consideração as características com redundância mínima e relevância máxima (BATTITI, 1994).

A seleção de uma característica remete à um problema de escolha de variáveis de entrada consideradas relevantes para a predição de um determinado resultado. Usando recursos relevantes, algoritmos de classificação podem melhorar sua precisão de previsão, encurtar o período de aprendizagem e resultar em conceitos mais simples, melhorando a eficácia e domínio da interpretação de um modelo de inferência (LIU; SETIONO, 1995).

O método *F-Score* (*Fisher Score*) para seleção de característica disponibiliza uma medida dada pela distância entre as médias das distribuições de duas classes (C_1 e C_2) em relação às suas variâncias Duda, Hart e Stork (2001). Quanto maior é o valor da classe calculado pelo *F-Score*, mais discriminativa e relevante é a característica para o experimento (GU; LI; HAN, 2012). O cálculo do *F-Score* é dado pela Equação 2.1

$$F(g) = \frac{\sum_{k=1}^{n} n_k(\mu_k^j - \mu^j)^2}{(\sigma^j)^2} \tag{2.1}$$

Em que:

- $(\sigma^j)^2 = \sum_{i=1}^{n} n_k(\sigma_k^j)^2$
- *F(g)* é a função que calcula o valor *F-score* para a característica g;

O Coeficiente de Correlação de *Pearson* para seleção de característica, mede o grau de relação da distribuição de duas classes. O do Coeficiente de Correlação de *Pearson* é definido pela Equação 2.2 (DUDA; HART; STORK, 2001):

$$\rho_j = \frac{\frac{1}{n-1}\sum_{i=0}^{n}(x_{ij} - \bar{x}) \times (y_{ij} - \bar{y})}{\sigma_{xj} \times \sigma_y} \tag{2.2}$$

Em que

- -1 $\leq \rho_j \leq$ 1;
- Se $\rho_j = 1$, existe total relação positiva entre as distribuições;
- Se $\rho_j = -1$, existe total relação negativa entre as distribuições;
- Se $\rho_j = 0$, as distribuições não possuem relação.

Os métodos de classificação podem ser agrupados em duas grandes categorias (DUDA; HART; STORK, 2001): métodos supervisionados e métodos não-supervisionados. Nos métodos da classificação supervisionada, as classes são previamente definidas pelo analista, isto é, definidas ou caracterizadas através das amostras de treinamento. Os métodos não-supervisionados oferecem um outro tipo de abordagem. Em alguns casos, existem problemas na área de reconhecimento de padrões, nas quais a natureza (ou definição) das classes, e mesmo o número de classes presentes são desconhecidos. Neste caso, o problema a ser tratado consiste não somente na classificação propriamente dita,

mas também na própria definição das classes. Ao contrário do método supervisionado, onde se tem um conhecimento prévio das classes, os métodos não-supervisionados atribuem à técnica ou ao algoritmo escolhido a tarefa de identificar as classes existentes num conjunto de dados (ANDREOLA, 2009).

Os SVMs (Support Vector Machine) foram introduzidos recentemente como uma técnica para resolver problemas de reconhecimento de padrões. Esta estratégia de aprendizagem foi proposta por Vapnik (VAPNIK; VAPNIK, 1998) e tem atraído a atenção dos pesquisadores devido as suas principais características, que são a sua boa capacidade de generalização e robustez diante de dados de grande dimensão (BARANOSKI, 2005).

Os algoritmos de aprendizagem de máquina SVM, têm como objetivo a determinação de limites de decisão que produzam uma separação ótima entre classes por meio da minimização dos erros Vapnik e Vapnik (1998).

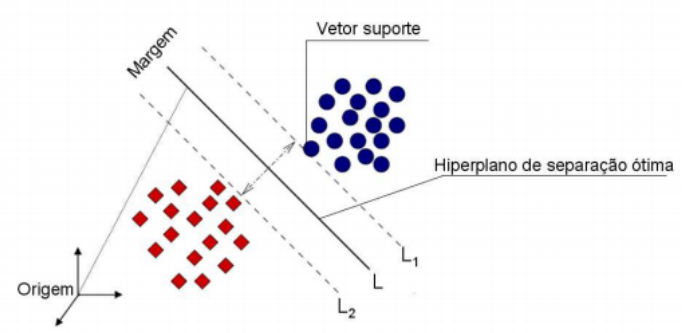

Figura 4 – Distribuição do Hiperplano
Fonte:(NASCIMENTO et al., 2009)

A separação ótima entre classes ocorre por meio de um hiperplano condicional (L) (Figura 4), tal que este plano é orientado para maximizar a margem (distância entre as bordas L1 e L2) e pelo ponto mais próximo de cada classe (NASCIMENTO et al., 2009).

O SVM possui quatro funções, sendo elas (i) linear, (ii) quadrática, (iii) polinomial e (iv) função de base radial (NASCIMENTO et al., 2009).

As funçoes *kernel* é o que diferenciam um modelo de SVM de outro, ou seja, a função realiza o mapeamento dos dados (HORTA et al., 2011).

3 Material e Métodos

O modelo da metodologia utilizada neste trabalho é apresentado na Figura 5.

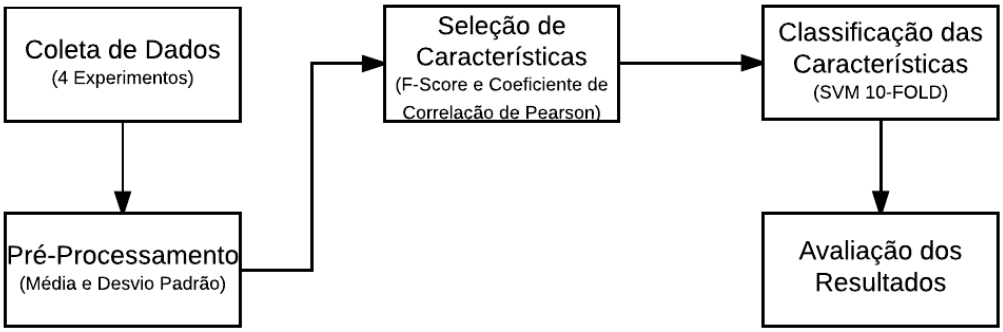

Figura 5 – Metodologia Proposta

Esse modelo representa cada etapa descrita nas 5 seções deste capítulo.

Para o desenvolvimento do trabalho, foi montada uma LAN com 7 computadores no laboratório 38 do prédio de Sistemas de Informação da UFVJM, 6 deles desktops de configurações iguais da marca DELL, compostos por processador Intel(R) Core(TM) i7 2.93GHz, 8GB de memória RAM DDR3, 500Gb de Disco Rígido e SO Ubuntu versão 15.04, o último sendo um notebook da marca HP, processador Intel(R) Core(TM) i3 2.53GHz, 4GB de memória RAM DDR2, 320Gb de Disco Rígido e SO Kali Linux versão 2017.1. Foi criada uma vlan 3028 para os testes, com porta única conectada ao roteador que gerenciava os IP's das máquinas locais ligadas ao *switch*.

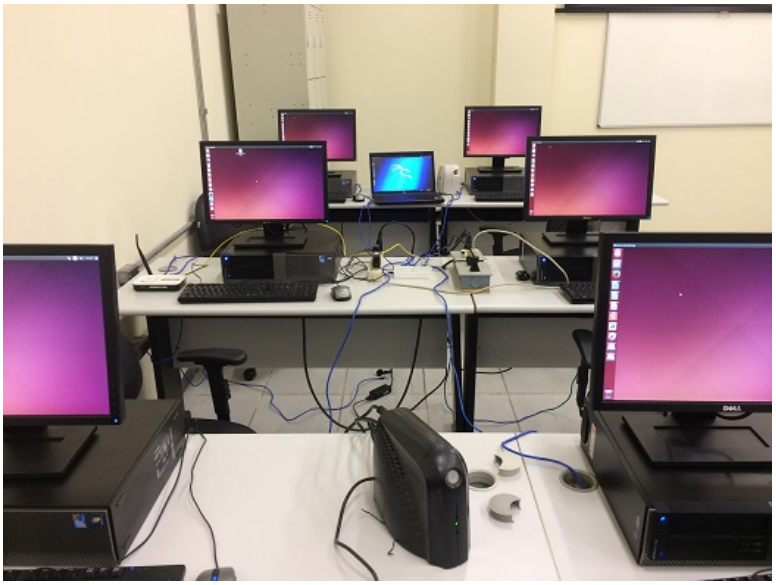

Figura 6 – LAN - *testebed*

A rede LAN foi sobrecarregada localmente, através de dois comandos do Linux utilizados em todos os computadores DELL: o **wget** e o **scp**. O **wget** foi utilizado como script *Bash*, para fazer 5 downloads simultâneos de uma ISO do Ubuntu 16.04 em cada máquina, em segundo plano. O comando **scp**, foi utilizado a cada par de máquinas, onde ocorria a transferência local de mais de 120Gb de arquivos através do protocolo SSH.

```
root@lambari06:/home/decom/Downloads# ./download.sh
--2017-08-08 14:33:17--  http://releases.ubuntu.com/xenial/ubuntu-16.04.3-desktop-amd64.iso
Resolvendo releases.ubuntu.com (releases.ubuntu.com)... 91.189.88.23, 2001:67c:1360:8001::26
Conectando-se a releases.ubuntu.com (releases.ubuntu.com)|91.189.88.23|:80... conectado.
A requisição HTTP foi enviada, aguardando resposta... 200 OK
Tamanho: 1587609600 (1,5G) [application/x-iso9660-image]
Salvando em: “ubuntu-16.04.3-desktop-amd64.iso.9”

 0% [                                                                                                                                  ]
```

Figura 7 – Comando **wget**

```
root@lambari06:/home/decom/Documentos# scp -r Arquivos/ decom@192.168.1.103:/home/decom
decom@192.168.1.103's password:
Kohavi-IJCAI-95.pdf                                                                                  100%  208KB
avaliacao.pdf                                                                                        100%  190KB
AULA03_RFON_TAE_BEHAVIOR_2.pdf                                                                       100%  781KB
0711323_09_cap_05.pdf                                                                                100% 1729KB
avalia.pdf                                                                                           100% 1692KB
AvaliacaoClassificadores.ppt                                                                         100%  302KB
CleberOlivoMSC.pdf                                                                                   100% 1063KB
DISSERTAÇÃO José Júnior de Oliveira Silva.pdf                                                        100% 7342KB   7
versao_final_revisada_depositada.pdf                                                                 100% 2027KB
v1n2a06.pdf                                                                                          100%  437KB
0711323_09_pretextual.pdf                                                                            100%   69KB
```

Figura 8 – Comando **scp**

O ataque DNS *Spoofing* foi realizado através do notebook HP com SO Kali Linux, utilizando a ferramenta Ettercap, que já vem instalada no sistema. O *Ettercap* possui 4 tipos de interface de usuário: somente texto, curses, GTK e daemon. Neste trabalho, foi usada a interface somente texto.

Antes de abrir o *Ettercap* para utilizar o comando de plugin do *Spoofing*, foi necessário modificar o arquivo etter.dns no diretório da ferramenta, para direcionar a página de configuração do ataque, para a máquina do atacante. A página do Facebook foi escolhida para o ataque, e direcionada para o IP da máquina com o Kali Linux e o *Ettercap*.

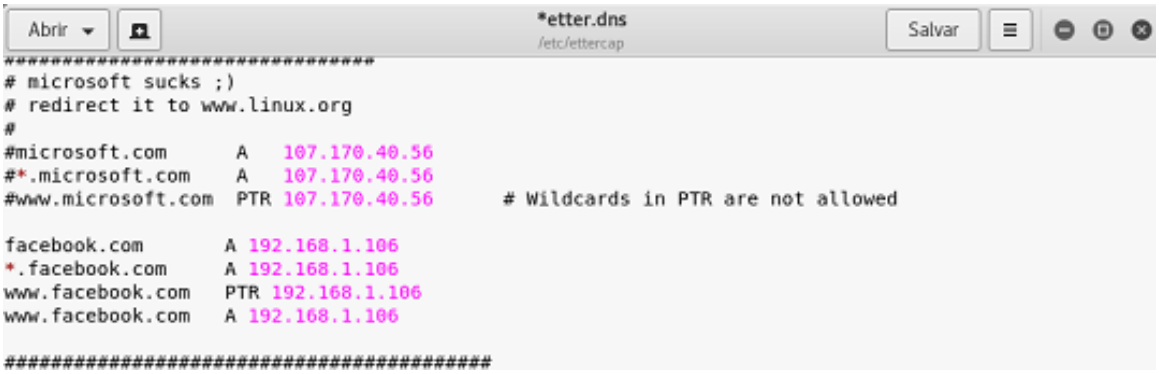

Figura 9 – Arquivo etter.dns

A ferramenta *Ettercap* possui o plugin de DNS *Spoofing*, executado através do comando **-P dns_spoof** pelo modo de interface somente texto no prompt de comando do Kali Linux **ettercap -T**. Compõe o comando: modo silencioso **-q**, tipo de redirecionamento do DNS **-M arp** e a interface de rede utilizada **-i eth0**. Após a execução, os hosts da LAN foram identificados e o plugin ativado. As requisições realizadas nos computadores da rede local para o site do Facebook, foram então redirecionadas para o computador atacante.

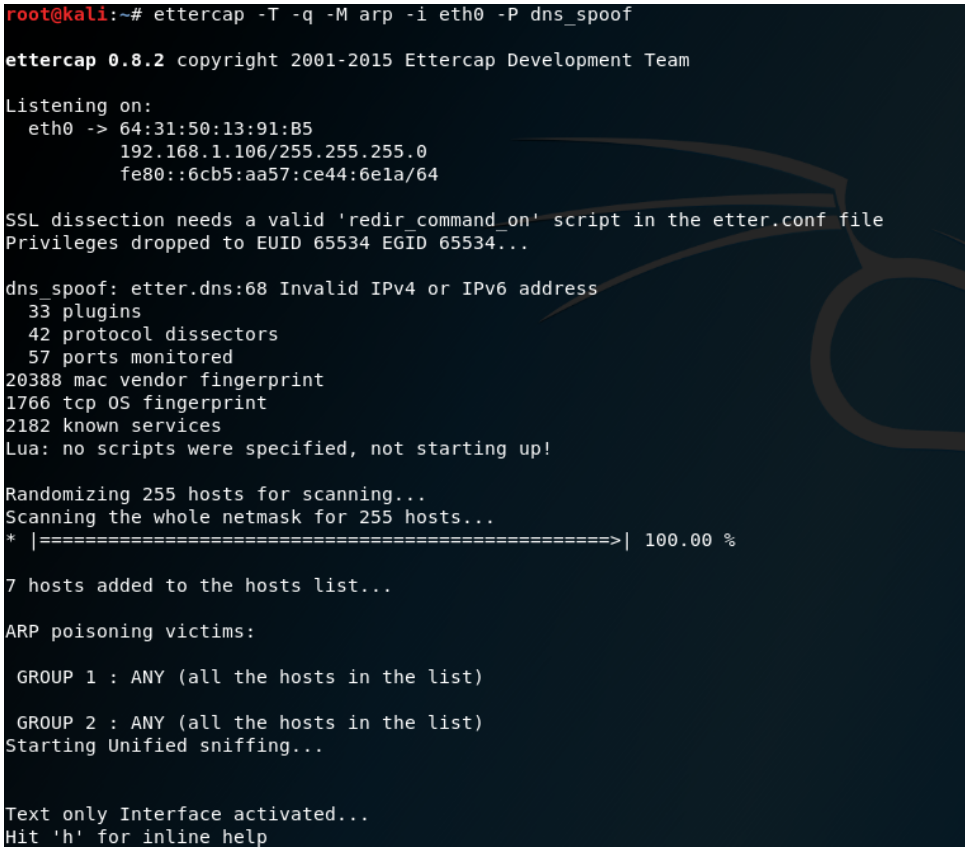

Figura 10 – *Ettercap* - DNS *Spoofing*

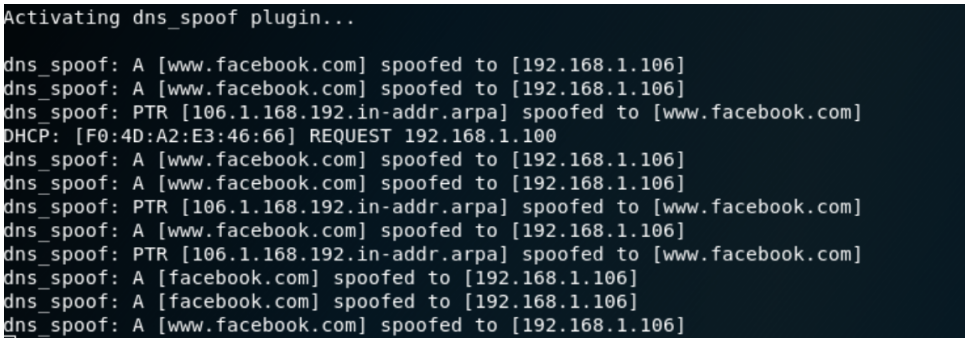

Figura 11 – *Ettercap* - Requisições ao site Facebook

Após a simulação do ataque com sucesso, a página de acesso ao Facebook ficou indisponível, devido ao redirecionamento à máquina atacante. Os comandos **ping** e **traceroute**, também confirmaram a eficácia do ataque, retornando o endereço IP do host atacante no lugar do verdadeiro endereço do site.

Figura 12 – Página indisponível

```
root@decom-OptiPlex-980:/home/decom/Downloads# ping www.facebook.com
PING www.facebook.com (192.168.1.106) 56(84) bytes of data.
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=1 ttl=64 time=0.245 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=2 ttl=64 time=0.280 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=3 ttl=64 time=0.524 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=4 ttl=64 time=0.249 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=5 ttl=64 time=0.489 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=6 ttl=64 time=0.247 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=7 ttl=64 time=0.225 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=8 ttl=64 time=0.359 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=9 ttl=64 time=0.436 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=10 ttl=64 time=0.289 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=11 ttl=64 time=0.236 ms
64 bytes from www.facebook.com (192.168.1.106): icmp_seq=12 ttl=64 time=0.332 ms
^C
--- www.facebook.com ping statistics ---
12 packets transmitted, 12 received, 0% packet loss, time 11001ms
rtt min/avg/max/mdev = 0.225/0.325/0.524/0.102 ms
root@decom-OptiPlex-980:/home/decom/Downloads#
```

Figura 13 – Comando **ping**

```
root@decom-OptiPlex-980:/home/decom/Downloads# traceroute -I www.facebook.com
traceroute to www.facebook.com (192.168.1.106), 30 hops max, 60 byte packets
 1  www.facebook.com (192.168.1.106)  0.293 ms  0.294 ms  0.292 ms
root@decom-OptiPlex-980:/home/decom/Downloads#
```

Figura 14 – Comando **traceroute**

3.1 Coleta de Dados

Segundo o estudo de Carvalho e Pelli (2017), as características identificadas em uma simulação de ataque são:

- Número de saltos (TTL): é identificado através do comando **traceroute**, que contabiliza o número de saltos que um pacote demora para chegar ao destino, passando por cada roteador, servidor e retornando ao requisitante;

- Tempo de resposta (RTT): é identificado através do comando **ping**, que contabiliza o tempo que um pacote leva para chegar ao servidor destino e voltar a máquina solicitante;
- Erros: um erro é caracterizado pelo comando **ping** quando o pacote excede o tempo máximo permitido, atribuindo o valor 1 a característica, e 2000ms ao tempo de resposta;
- Velocidade: é obtida através do comando **wget** que faz o download de um arquivo, armazenando sua velocidade instantânea.

Baseadas nestas informações, as mesmas características foram utilizadas neste trabalho.

Quatro experimentos foram estipulados para a coleta e análise das características:

- Experimento 1: nenhuma sobrecarga;
- Experimento 2: sobrecarga local através do comando **scp**;
- Experimento 3: sobrecarga local através do comando **wget**;
- Experimento 4: sobrecarga local através dos dois comandos, **scp** e **wget**;

Os dados foram coletados na rede local criada para a simulação do ataque e as sobrecargas de rede, por meio de um script *Bash* desenvolvido para armazenar os valores das características em uma planilha. A coleta ocorreu durante uma semana, no horário de 14:00h às 22:00h. O script era iniciado em dois momentos: quando ocorria o ataque de DNS *Spoofing*, e sem a ocorrência deste. Para separar os dois momentos, mais uma característica foi atribuída: a variável classe, que assumiu o valor 1 para o *Spoofing*, e o valor 2 quando não havia ataque.

Os computadores com o Ubuntu executaram o script ao mesmo tempo, e armazenaram 500 dados de cada característica, conforme a classe de ataque, nos 4 experimentos. Ao final da coleta, cada experimento contava com 6.000 dados de cada característica (3.000 de cada classe). Ao todo, foram coletados 24.000 dados, dos 4 experimentos.

Tabela 1 – Análise descritiva dos dados

Variável	Mínimo	Máximo	Média	Desvio Padrão
Velocidade(Kbps)	10	998	230.86	200.59
Tempo de Resposta (ms)	0.116	2.000	297.488	702.89
Número de Saltos (un)	1	30	6.02	6.49
Erros (un)	0	1	0.38	0.52

3.2 Pré-processamento dos Dados

Para obter dados mais homogêneos, as características passaram por um reagrupamento a cada 10 dados, que foram transformados em média e desvio padrão. Nesta etapa,

foi utilizado um script desenvolvido na linguagem R, que processava a planilha gerada pela coleta de dados, fornecendo novos valores para a base de dados.

A média e o desvio padrão foram escolhidos pois são a melhor estimativa de valor verdadeiro, dentre n quantidades de x "determinações". Este pré-processamento foi necessário para diminuir erros quanto a avaliação, tornando-a mais precisa.

O pré-processamento, deu origem a uma nova base de dados para cada experimento com 600 novos valores de 8 características: Média de Velocidade, Desvio Padrão de Velocidade, Média do Tempo de Resposta, Desvio Padrão do Tempo de Resposta, Média de Saltos, Desvio Padrão de Saltos, Média de Erros e Desvio Padrão de Erros.

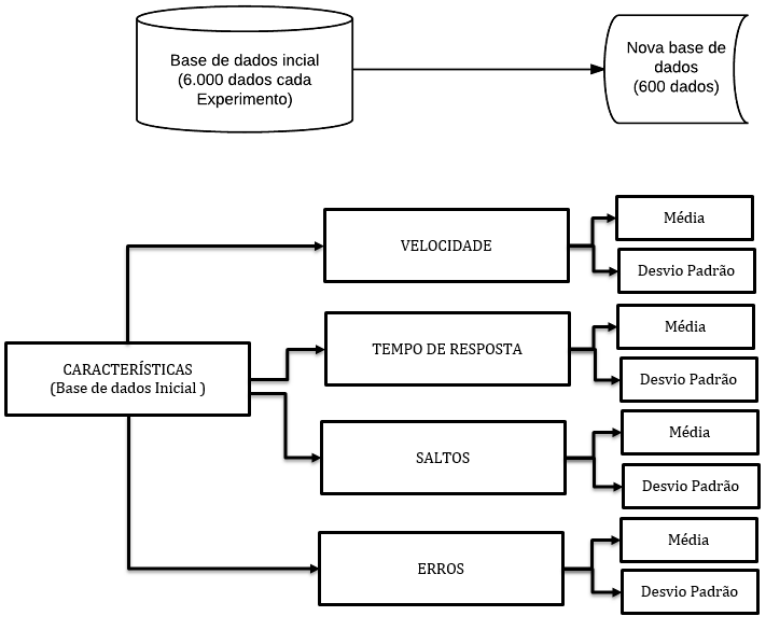

Figura 15 – Etapas do pré-processamento

3.3 Seleção de Características

O pré-processamento de dados permitiu diminuir a dispersão de erros nas características, que foram ranqueadas através das técnicas de seleções conhecidas: *F-Score* e Coeficiente de correlação de *Pearson*.

As técnicas de seleção foram implementadas em scripts na linguagem R, que processavam a nova base de dados de cada experimento. Os resultados foram valores representativos, que identificavam quais eram as características mais relevantes para se aplicar o classificador. Na Tabela do *F-Score*, quanto maior o valor da característica, mais discriminativa e relevante ela é para o estudo. Já o Coeficiente de correlação de

Pearson, assume valores entre -1(correlação negativa) e 1(correlação positiva), medindo a intensidade da relação das duas classes.

Os gráficos gerados pelo *F-Score*, foram assimilados aos resultados das Tabelas e foi realizada uma análise visual da distância das distribuições das duas classes em relação às suas variâncias. Nos gráficos, ficou estipulado a cor vermelha para a classe 1 (*Spoofing*) e a cor azul para a classe 2 (sem ataque).

3.4 Classificação de Características

O classificador escolhido para este trabalho foi o SVM, pois apresenta uma boa generalização, sendo utilizado em diversas áreas de conhecimento. O SVM foi implementado em um script na linguagem R, que recebeu como dados de treinamento e teste, as características mais relevantes obtidas através das técnicas de seleção.

Para o aprendizado supervisionado, o tamanho passado para o treinamento foi de 70% da base de dados, onde eram fornecidos os índices que indicavam a classe em que as características se encontravam. Os outros 30% utilizados como teste, não possuía a informação da classe. O treinamento é a etapa que provê conhecimento para o SVM classificar as características sem informação da base de teste.

O SVM foi calibrado para utilzar os melhores parâmetros. Após o pré-processamento de verificação, foram definidos: Kernel Radial, cost=10, e gamma=0.5.

O método de validação cruzada (10-fold *fold cross-validation*), foi utilizado como forma de evitar a superposição dos dados de teste: a base de dados foi dividida em 10 subconjuntos, cada subconjunto aleatório foi utilizado como dado de teste, enquanto os demais eram submetidos ao treinamento. Esse procedimento foi repetido 10 vezes. Ao final das 10 repetições, a matriz de confusão era gerada e foi possível calcular a acurácia (% das características que foram classificados corretamente) e precisão (% com que o classificador fornece resultados semelhantes após várias repetições) do classificador SVM.

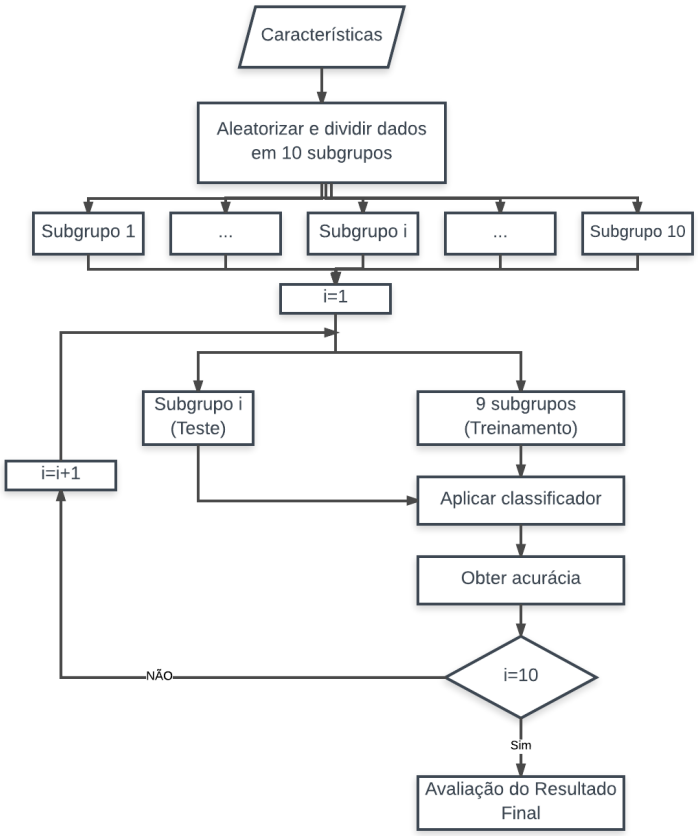

Figura 16 – Processos da classificação de características

3.5 Avaliação dos Resultados

Para comparar os 4 experimentos, foram utilizadas análises dos gráficos de dispersão e das matrizes de confusão geradas pelo SVM, assim como dos valores de acurácia e precisão encontrados.

Um exemplo da matriz de confusão, é apresentado na Figura 17:

		PREDITO	
		Classe 1	**Classe 2**
VERDADEIRO	**Classe 1**	VP	FN
	Classe 2	FP	VN

Figura 17 – Matriz de confusão
Fonte: Adaptado de (NONATO; OLIVEIRA, 2013)

Em que:

VP: Verdadeiro Positivo;

VN: Verdadeiro Negativo;

FP: Falso Positivo;

FN: Falso Negativo;

A acurácia foi calculada pela equacão:

$$Acurácia = \frac{VP + VN}{VP + VN + FN + FP} \tag{3.1}$$

E a precisão pela equação:

$$Precisão = \frac{VP}{VP + FP} \tag{3.2}$$

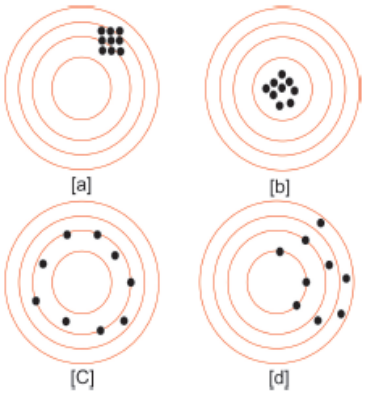

Figura 18 – Representação de Acurácia e Precisão
Fonte: (SILVA; SPATTI; FLAUZINO, 2010)

A Figura 18 apresenta o conceito de acurácia e precisão: No alvo [a], os dados não são acurados devido à presença de tendência central no alvo (boa precisão); no alvo [b] os dados são acurados e precisos; no alvo [c] os dados são pouco acurados e pouco precisos, no alvo [d] os dados não são precisos e são tendenciosos, portanto, não são acurados (SILVA; SPATTI; FLAUZINO, 2010).

4 Resultados e Discussão

4.1 Seleção de Características

Os resultados obtidos pela aplicação das técnicas de seleção *F-Score* e Coeficiente de correlação de *Pearson*, para cada experimento, podem ser vistos nas Tabela 3 e 4.

Tabela 2 – Tabela de Referência das Características

Código	Nome
1	Média da Velocidade
2	Desvio Padrão da Velocidade
3	Média do Tempo de Resposta
4	Desvio Padrão do Tempo de Resposta
5	Média de Saltos
6	Desvio Padrão de Saltos
7	Média de Erros
8	Desvio Padrão de Erros

Tabela 3 – Tabela de Ranqueamento - F-Score

EXPERIMENTO 1		EXPERIMENTO 2		EXPERIMENTO 3		EXPERIMENTO 4	
CÓD	F-SCORE	CÓD	F-SCORE	CÓD	F-SCORE	CÓD	F-SCORE
5	223.1472	5	31.0376	5	37.1122	5	63.9593
4	17.2143	8	8.2287	8	15.5828	8	29.8278
8	17.1937	4	8.2187	4	15.4549	4	29.4948
3	2.4385	3	2.6374	1	3.2946	3	3.7505
7	2.3144	7	1.9709	3	2.9592	7	3.4425
1	0.7642	1	0.6502	7	2.5204	6	0.8264
2	0.6826	6	0.6212	2	1.5635	2	0.0979
6	0.5251	2	0.5332	6	1.275	1	0.0129

Tabela 4 – Tabela de Ranqueamento - Coeficiente de Correlação de Pearson

EXPERIMENTO 1		EXPERIMENTO 2		EXPERIMENTO 3		EXPERIMENTO 4	
CÓD	PEARSON	CÓD	PEARSON	CÓD	PEARSON	CÓD	PEARSON
5	0.9977	5	0.9843	5	0.9868	5	0.9922
4	0.9723	8	0.9444	8	0.9694	8	0.9837
8	0.9722	4	0.9443	4	0.9692	4	0.9835
3	0.8425	3	0.8519	1	-0.8762	3	0.8888
7	0.8360	7	0.8149	3	0.8649	7	0.8806
1	0.6587	1	0.6283	7	0.8465	6	0.6732
2	0.6375	6	0.6196	2	-0.7814	2	0.2990
6	0.5874	2	0.5903	6	0.7491	1	-0.1132

Ao comparar os resultados dos 4 experimentos, observou-se que nas duas métricas de seleção, as características Média de Saltos, Desvio Padrão do Tempo de Resposta, Desvio Padrão de Erros e Média do Tempo de Resposta foram classificadas como mais relevantes.

Somente no experimento 3, a característica Média de Velocidade aparece como a terceira mais relevante pelo método de seleção *F-Score*, mas assume correlação negativa pela seleção de Pearson. Por se tratar de uma exceção, a característica não foi relacionada para aplicação do SVM, sendo necessário novos experimentos para analisar o comportamento da rede local.

4.1.1 Gráficos de dispersão

Os Gráficos de dispersão referente as características mais relevantes selecionadas pelo ranqueamento dos métodos de seleção, são apresentados nas Figuras 19, 20 e 21, onde a classe que representa o ataque DNS *Spoofing* foi representada pela cor vermelha e a classe que não configura ataque pela cor azul.

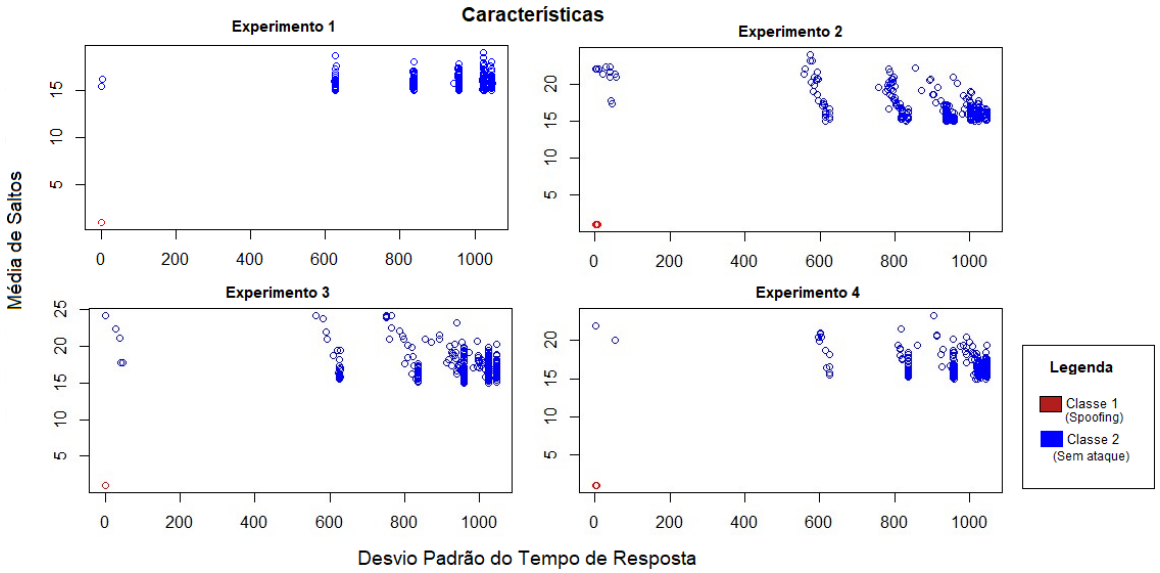

Figura 19 – Dispersão da Média de Saltos X Desvio Padrão do Tempo de resposta para cada experimento

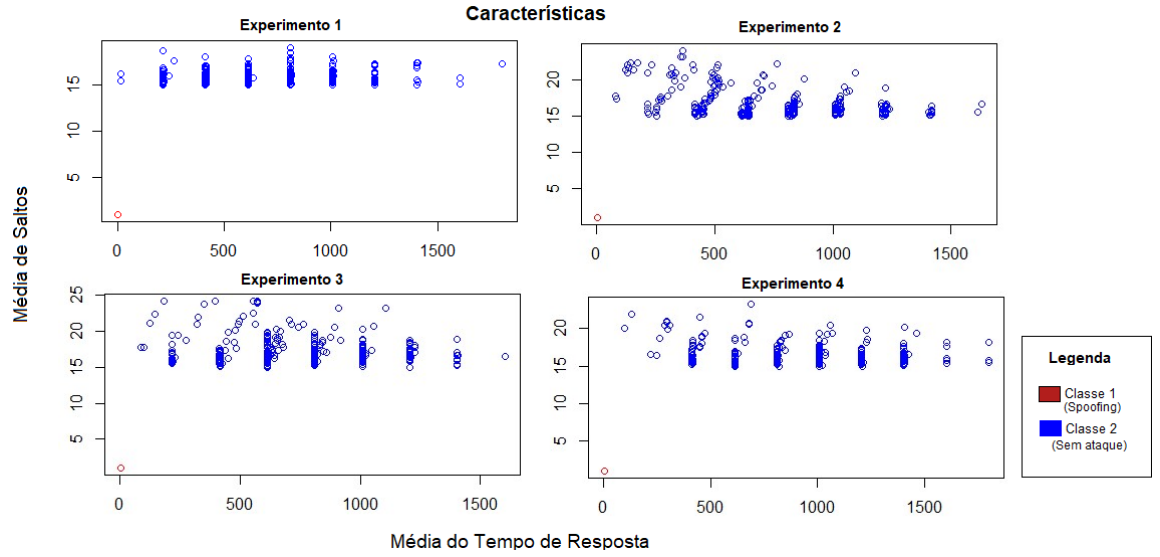

Figura 20 – Dispersão da Média de Saltos X Média do Tempo de resposta para cada experimento

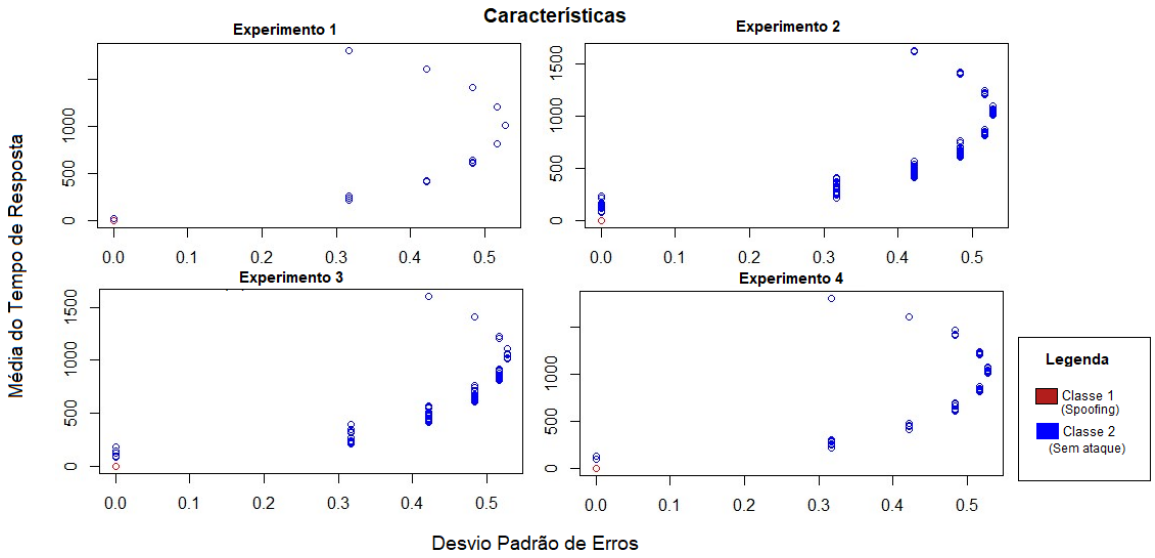

Figura 21 – Dispersão da Média do Tempo de Resposta X Desvio Padrão de Erros para cada experimento

Observou-se pelos Gráficos de dispersão, comparados os quatro experimentos, que as características selecionadas pelas métricas de seleção como mais relevantes não apresentaram alterações significativas, na relação entre a distribuição e as variânicias das classes determinadas pelo ataque/não ataque.

No resultado referente a dispersão das características Média do Tempo de Reposta e Desvio Padrão de Erros, sugere-se a existência de uma correlação entre as duas características. Ambas são identificadas pelo comando **ping**, e a assimilação de um erro afeta diretamente o Tempo de Resposta.

O resultado da comparação da característica Média de Saltos, nos quatro experimentos, evidencia o valor discriminativo da característica em relação às classes de ataque/não ataque.

4.2 Classificação de Características

Esta seção é dividida em 4 subseções, onde são apresentados e discutidos os resultados da aplicação do classificador SVM para cada característica selecionada como relevante pelas técnicas de seleção.

4.2.1 Resultados do SVM para a característica Média do Tempo de Resposta

Tabela 5 – Matriz média de confusão, acurácia e precisão da Média do Tempo de Resposta

Experimento	Acurácia Média	Precisão Média	VP	FN	FP	VN
1	99,40% ± 0,44%	98,81%± 0,85%	87	0	1	86
2	99,25% ± 0,88%	98,55% ± 1,68%	87	0	2	85
3	99,40% ± 0,47%	98,81% ± 0,93%	87	0	1	86
4	99,71% ± 0,40%	99,43% ± 0,77%	87	0	0	87

Na Tabela 5 é apresentada a acurácia média, precisão média e matriz média de confusão, resultados da aplicação do SVM para a característica Média do Tempo de Resposta.

As taxas de acerto do classificador, para todos os experimentos é em média 99,25% ± 0,88%, o que caracteriza a característica Média do Tempo de Resposta como altamente relevante para a identificação de ataques *Spoofing*. Através da análise da matriz média de confusão, observou-se que o classificador SVM conseguiu indentificar todas os dados da classe de ataque corretamente, os erros são caracterizados pelos falsos positivos.

A Média do Tempo de Resposta em ataques DNS *Spoofing* é muito baixa, pois os computadores encontram-se na mesma rede local. Devido ao cálculo da média, alguns valores normais podem se aproximar daqueles caracterizados pelo ataque, resultando nos falsos positivos encontrado pelo SVM.

4.2.2 Resultados do SVM para a característica Desvio Padrão do Tempo de Resposta

Tabela 6 – Matriz média de confusão, acurácia e precisão do Desvio Padrão do Tempo de Resposta

Experimento	Acurácia Média	Precisão Média	VP	FN	FP	VN
1	99,57% ± 0,41%	99,15%± 0,81%	87	0	1	86
2	98,16% ± 0,61%	96,47% ± 1,14%	87	0	4	83
3	99,19% ± 0,63%	98,43% ± 1,22%	87	0	2	85
4	99,60% ± 0,33%	99,21% ± 0,65%	87	0	1	86

Na Tabela 6 é apresentada a acurácia média, precisão média e matriz média de confusão, resultados da aplicação do SVM para a característica Desvio Padrão do Tempo de Resposta.

A análise dos resultados, pode ser feita de maneira similar a característica Média do Tempo de Resposta, porém os falsos positivos da matriz média de confusão estão relacionados ao cálculo do desvio padrão.

4.2.3 Resultados do SVM para a característica Média de Saltos

Tabela 7 – Matriz média de confusão, acurácia e precisão da Média de Saltos

Experimento	Acurácia Média	Precisão Média	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	100%	100%	87	0	0	87
3	100%	100%	87	0	0	87
4	100%	100%	87	0	0	87

Na Tabela 7 é apresentada a acurácia média, precisão média e matriz média de confusão, resultados da aplicação do SVM para a característica Média de Saltos.

A característica Média de Saltos, é a característica conhecida mais discriminante e relevante na indentificação de ataques DNS *Spoofing*. Na metodologia deste trabalho, encontra-se um exemplo (Figura 14) que justifica esta afirmação.

O número de saltos em um ataque DNS *Spoofing* sempre vai ser 1, enquanto em requisições normais, depende do caminho realizado pelo pacote até o destino.

4.2.4 Resultados do SVM para a característica Desvio Padrão de Erros

Tabela 8 – Matriz média de confusão, acurácia e precisão do Desvio Padrão de Erros

Experimento	Acurácia Média	Precisão Média	VP	FN	FP	VN
1	99,45% ± 0,59%	98,87% ± 1,14%	87	0	1	86
2	97,84% ± 0,85%	95,89% ± 1,55%	87	0	5	82
3	99,10% ± 0,66%	98,26% ± 1,27%	87	0	2	85
4	99,62% ± 0,33%	99,26% ± 0,66%	87	0	1	86

Na Tabela 8 é apresentada a acurácia média, precisão média e matriz média de confusão, resultados da aplicação do SVM para a característica Desvio Padrão de Erros.

As taxa de acerto do classificador, para todos os experimentos é em média 97,84% ± 0,85%, confirmando a relevância da característica apurada pelas técnicas de seleção. Através da matriz média de confusão, observou-se que o classificador é capaz de identificar corretamente todos os dados da classe de ataque. Os erros, classificados como falsos positivos, significam que o classificador errou em atribuir uma característica da classe sem ataque.

Em um ataque DNS *Spoofing*, não há como atribuir a característica erro, pois o pacote não excede o tempo limite. Porém, quando a rede não está sob ataque, ela também pode não apresentar erros, por isso alguns dados de Desvio Padrão de Erros são classificados como falsos positivos.

5 Conclusão

As métricas de avaliação dos resultados, *F-Score*, Coeficiente de Correlação de *Pearson*, Gráfico de dispersão e Matriz de confusão, são complementares, e podem ser utilizadas para a análise consistente dos dados.

Foi observado que as alterações nas cargas da rede local não afetam de forma significativa as características relevantes na identificação de ataques DNS *Spoofing* pelas técnicas de reconhecimento de padrões.

As estimativas de aprendizagem geradas pelo classificador SVM se apresentam semelhantes, com alta acurácia de predição de novos ataques, através da seleção das características ranqueadas. Através da matriz média de confusão, pôde se perceber que para as características Média de Saltos, Média do Tempo de Resposta e Desvio Padrão de Erros, o SVM sempre classifica corretamente os dados da classe *Spoofing* (os erros identificados são falsos positivos).

6 Trabalhos Futuros

Para o desenvolvimento de trabalhos futuros, verificou-se a necessidade de repetir os experimentos, analisando o comportamento da rede local.

Como proposta para futuros projetos, sugere-se: a simulação de outras cargas na rede local por diferentes métodos; dividir e analisar os experimentos em diferentes cargas estimadas pela porcentagem; utilizar uma rede local maior, para coleta e simulação da sobrecarga da rede; desenvolver uma aplicação com as características mais relevantes e avaliar a detecção do *Spoofing* .

Referências

AL.HAJERI, A. Dns spoofing attack support of the cyber defense initiative. *SANS Institute*, n. v2.1, 2000–2002. Disponível em: <https://www.giac.org/paper/gcih/364/dns-spoofing-attack/103863>. Citado 3 vezes nas páginas 16, 17 e 18.

ANDREOLA, R. Support vector machines na classificação de imagens hiperespectrais. Dissertação submetida ao Programa de PósGraduação em Sensoriamento Remoto do Centro Estadual de Pesquisas em Sensoriamento Remoto e Metereologia UFRGS. 2009. Disponível em: <http://www.ufrgs.br/srm/ppgsr/publicacoes/Dissert_RafaelaAndreola.pdf>. Citado na página 20.

BARANOSKI, F. Verificação da autoria em documentos manuscritos usando o svm. 2005. Citado 2 vezes nas páginas 18 e 20.

BATTITI, R. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, v. 5, p. 537–550, 1994. Citado na página 19.

CARVALHO, H. C. F. B.; PELLI, E. Técnicas de reconhecimento de padrões para identificação de ataques de dns. *Revista Brasileira de Computação Aplicada*, v. 9, p. 99–110, Julho 2017. ISSN 2176-6649. Citado 2 vezes nas páginas 12 e 25.

COMER, D. *Redes de computadores e Internet: abrange transmissão de dados, ligação inter-redes e Web*. 2ª. ed. [S.l.]: Bookman, 2001. Citado na página 15.

COSTA, D. *DNS - Um Guia para Administradores de Redes*. Rio de Janeiro: BRASPORT, 2007. ISBN 9788574522920. Citado na página 16.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. New York: Wiley, 2001. Citado 2 vezes nas páginas 18 e 19.

FOROUZAN, B. *Comunicação de Dados e Redes de Computadores*. 4ª. ed. [S.l.]: McGraw Hill Brasil, 2009. Citado na página 14.

GU, Q.; LI, Z.; HAN, J. Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*, 2012. Citado na página 19.

HORTA, E. G. et al. Extração de características e casamento de padrões aplicados à estimação de posição de um VANT. *UFMG*, 2011. Citado na página 20.

KUROSE, J.; ROSS, K. *Redes de computadores e a internet: uma abordagem top-down*. 5ª. ed. [S.l.: s.n.], 2010. ISBN 9788588639973. Citado 2 vezes nas páginas 12 e 14.

LIU, H.; SETIONO, R. Feature selection and discretization of numeric attributes. In: *In Proceedings of the Seventh International Conference on Tools with Artificial Intelligence*. [S.l.: s.n.], 1995. p. 388–391. Citado na página 19.

MOHAMMED, A. H. et al. Dns protection against spoofing and poisoning attacks. *International Conference on Information Science and Control Engineering, IEE Computer Society*, v. 12, p. 1308–1312, 2016. Citado na página 12.

NASCIMENTO, R. F. F. et al. O algoritmo support vector machines (svm): avaliação da separação ótima de classes em imagens ccd-cbers-2. *Simpósio Brasileiro de Sensoriamento Remoto*, v. 14, p. 2079–2086, 2009. Citado na página 20.

NONATO, R. T.; OLIVEIRA, S. R. d. M. Técnicas de mineração de dados para identificação de áreas com cana-de-açúcar em imagens landsat 5. *Engenharia Agrícola*, scielo, v. 33, p. 1268 – 1280, 12 2013. ISSN 0100-6916. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-69162013000600019&nrm=iso>. Citado na página 29.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. *Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas Curso Prático*. Artliber editora ltda. [S.l.: s.n.], 2010. ISBN 9788588098534. Citado na página 30.

SYMANTECCORPORATION. *Norton Cyber Security Insights Report 2016*. 2016. Site. Disponível em: <https://www.symantec.com/content/dam/symantec/br/docs/reports/2016-norton-cyber-security-insights-comparisons-brazil-pt.pdf>. Acesso em: 11 mar. 2017. Citado na página 12.

TANENBAUM, A. *Redes de computadores*. 4ª. ed. Rio de Janeiro: Elsevier Brasil, 2003. Citado na página 14.

TOU, J. T.; GONZALES, R. C. *Pattern Recognition Principles*. New York: Addison-Wesley, 1974. 377 p. Citado na página 12.

VAPNIK, V. N.; VAPNIK, V. *Statistical learning theory*. New York: Wiley New York, 1998. v. 1. Citado na página 20.

WEARESOCIAL. *Digital in 2017: Global Overview*. 2017. Site. Disponível em: <https://wearesocial.com/uk/special-reports/digital-in-2017-global-overview>. Acesso em: 23 mar. 2017. Citado na página 12.

WRIGHTSON, T. *Segurança de Redes Sem Fio: Guia do Iniciante*. Porto Alegre: Bookman Editora, 2014. Citado na página 18.

APÊNDICE A – Resultados SVM

A.1 Característica Média de Saltos

Tabela 9 – Experimento 1 - Média de Saltos

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	100%	100%	87	0	0	87
3	100%	100%	87	0	0	87
4	100%	100%	87	0	0	87
5	100%	100%	87	0	0	87
6	100%	100%	87	0	0	87
7	100%	100%	87	0	0	87
8	100%	100%	87	0	0	87
9	100%	100%	87	0	0	87
10	100%	100%	87	0	0	87

Tabela 10 – Experimento 2 - Média de Saltos

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	100%	100%	87	0	0	87
3	100%	100%	87	0	0	87
4	100%	100%	87	0	0	87
5	100%	100%	87	0	0	87
6	100%	100%	87	0	0	87
7	100%	100%	87	0	0	87
8	100%	100%	87	0	0	87
9	100%	100%	87	0	0	87
10	100%	100%	87	0	0	87

Tabela 11 – Experimento 3 - Média de Saltos

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	100%	100%	87	0	0	87
3	100%	100%	87	0	0	87
4	100%	100%	87	0	0	87
5	100%	100%	87	0	0	87
6	100%	100%	87	0	0	87
7	100%	100%	87	0	0	87
8	100%	100%	87	0	0	87
9	100%	100%	87	0	0	87
10	100%	100%	87	0	0	87

Tabela 12 – Experimento 4 - Média de Saltos

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	100%	100%	87	0	0	87
3	100%	100%	87	0	0	87
4	100%	100%	87	0	0	87
5	100%	100%	87	0	0	87
6	100%	100%	87	0	0	87
7	100%	100%	87	0	0	87
8	100%	100%	87	0	0	87
9	100%	100%	87	0	0	87
10	100%	100%	87	0	0	87

A.2 Característica Desvio Padrão do Tempo de Resposta

Tabela 13 – Experimento 1 - Desvio Padrão do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	99,42%	98,86%	87	0	1	86
3	99,42%	98,86%	87	0	1	86
4	100%	100%	87	0	0	87
5	99,42%	98,86%	87	0	1	86
6	99,42%	98,86%	87	0	1	86
7	98,85%	97,75%	87	0	2	85
8	100%	97,75%	87	0	2	85
9	98,27%	96,66%	87	0	3	84
10	99,42%	98,86%	87	0	1	86

Tabela 14 – Experimento 2 - Desvio Padrão do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	97,12%	94,56%	87	0	5	82
2	98,85%	97,75%	87	0	2	85
3	97,70%	95,60%	87	0	4	83
4	98,27%	96,66%	87	0	3	84
5	97,70%	95,60%	87	0	4	83
6	99,42%	98,86%	87	0	1	86
7	97,12%	94,56%	87	0	5	82
8	98,85%	97,75%	87	0	2	85
9	97,12%	94,56%	87	0	5	82
10	100%	97,75%	87	0	2	85

Tabela 15 – Experimento 3 - Desvio Padrão do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	98,27%	96,66%	87	0	3	84
3	98,85%	97,75%	87	0	2	85
4	99,42%	98,86%	87	0	1	86
5	100%	100%	87	0	0	87
6	99,42%	98,86%	87	0	1	86
7	100%	100%	87	0	0	87
8	100%	100%	87	0	0	87
9	98,85%	97,75%	87	0	2	85
10	99,42%	98,86%	87	0	1	86

Tabela 16 – Experimento 4 - Desvio Padrão do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	100%	100%	87	0	0	87
3	99,42%	98,86%	87	0	1	86
4	100%	100%	87	0	0	87
5	100%	100%	87	0	0	87
6	99,42%	98,86%	87	0	1	86
7	100%	100%	87	0	0	87
8	99,42%	98,86%	87	0	1	86
9	100%	100%	87	0	0	87
10	98,85%	97,75%	87	0	2	85

A.3 Característica Desvio Padrão de Erros

Tabela 17 – Experimento 1 - Desvio Padrão de Erros

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	99,42%	98,86%	87	0	1	86
2	99,42%	98,86%	87	0	1	86
3	99,42%	98,86%	87	0	1	86
4	100%	100%	87	0	0	87
5	99,42%	98,86%	87	0	1	86
6	98,85%	97,75%	87	0	2	85
7	98,85%	97,75%	87	0	2	85
8	100%	100%	87	0	0	87
9	99,42%	98,86%	87	0	1	86
10	100%	100%	87	0	0	87

Tabela 18 – Experimento 2 - Desvio Padrão de Erros

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	97,70%	95,60%	87	0	4	83
2	98,27%	96,66%	87	0	3	84
3	96,55%	93,54%	87	0	6	81
4	95,97%	92,55%	87	0	7	80
5	98,85%	97,75%	87	0	2	85
6	99,42%	98,86%	87	0	1	86
7	98,27%	96,66%	87	0	3	84
8	98,27%	96,66%	87	0	3	84
9	98,27%	96,66%	87	0	3	84
10	99,42%	98,86%	87	0	1	86

Tabela 19 – Experimento 3 - Desvio Padrão de Erros

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	98,85%	97,75%	87	0	2	85
2	100%	100%	87	0	0	87
3	98,27%	96,66%	87	0	3	84
4	98,85%	97,75%	87	0	2	85
5	99,42%	98,86%	87	0	1	86
6	99,42%	98,86%	87	0	1	86
7	100%	100%	87	0	0	87
8	98,85%	97,75%	87	0	2	85
9	100%	100%	87	0	0	87
10	98,27%	96,66%	87	0	3	84

Tabela 20 – Experimento 4 - Desvio Padrão de Erros

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	100%	100%	87	0	0	87
2	100%	100%	87	0	0	87
3	100%	100%	87	0	0	87
4	100%	100%	87	0	0	87
5	99,42%	98,86%	87	0	1	86
6	100%	100%	87	0	0	87
7	100%	100%	87	0	0	87
8	98,85%	97,75%	87	0	2	85
9	99,42%	98,86%	87	0	1	86
10	99,42%	98,86%	87	0	1	86

A.4 Característica Média do Tempo de Resposta

Tabela 21 – Experimento 1 - Média do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	99,42%	98,86%	87	0	1	86
2	100%	100%	87	0	0	87
3	98,85%	97,75%	87	0	2	85
4	98,85%	97,75%	87	0	2	85
5	100%	100%	87	0	0	87
6	100%	100%	87	0	0	87
7	100%	100%	87	0	0	87
8	100%	100%	87	0	0	87
9	99,42%	98,86%	87	0	1	86
10	99,42%	98,86%	87	0	1	86

Tabela 22 – Experimento 2 - Média do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	99,42%	98,86%	87	0	1	86
2	98,85%	97,75%	87	0	2	85
3	98,85%	97,75%	87	0	2	85
4	99,42%	98,86%	87	0	1	86
5	100%	100%	87	0	0	87
6	99,42%	98,86%	87	0	1	86
7	98,85%	97,75%	87	0	2	85
8	98,27%	96,66%	87	0	3	84
9	98,85%	97,75%	87	0	2	85
10	98,85%	97,75%	87	0	2	85

Tabela 23 – Experimento 3 - Média do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	99,42%	98,86%	87	0	1	86
2	100%	100%	87	0	0	87
3	100%	100%	87	0	0	87
4	98,85%	97,75%	87	0	2	85
5	98,85%	97,75%	87	0	2	85
6	99,42%	98,86%	87	0	1	86
7	99,42%	98,86%	87	0	1	86
8	100%	100%	87	0	0	87
9	98,27%	96,66%	87	0	3	84
10	99,42%	98,86%	87	0	1	86

Tabela 24 – Experimento 4 - Média do Tempo de Resposta

Repetição	Acurácia	Precisão	VP	FN	FP	VN
1	98,85%	97,75%	87	0	2	85
2	99,42%	98,86%	87	0	1	86
3	99,42%	98,86%	87	0	1	86
4	98,85%	97,75%	87	0	2	85
5	100%	100%	87	0	0	87
6	99,42%	98,86%	87	0	1	86
7	99,42%	98,86%	87	0	1	86
8	100%	100%	87	0	0	87
9	98,85%	97,75%	87	0	2	85
10	98,85%	97,75%	87	0	2	85