SISTEMA DE DETECÇÃO E CLASSIFICAÇÃO DE INTRUSÃO EM REDES AD HOC UTILIZANDO REDES NEURAIS ARTIFICIAIS E ALGORITMO K-MÉDIAS

DANIEL ROSA CANÊDO

TESE DE DOUTORADO EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E AUTOMAÇÃO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

SISTEMA DE DETECÇÃO E CLASSIFICAÇÃO DE INTRUSÃO EM REDES AD HOC UTILIZANDO REDES NEURAIS ARTIFICIAIS E ALGORITMO K-MÉDIAS

DANIEL ROSA CANÊDO

Orientador: PROF. DR. ALEXANDRE RICARDO SOARES ROMARIZ, ENE/UNB

TESE DE DOUTORADO EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E AUTOMAÇÃO

PUBLICAÇÃO: PPGEA.TD - 141/2019
BRASÍLIA/DF: JUNHO - 2019

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

SISTEMA DE DETECÇÃO E CLASSIFICAÇÃO DE INTRUSÃO EM REDES AD HOC UTILIZANDO REDES NEURAIS ARTIFICIAIS E ALGORITMO K-MÉDIAS

DANIEL ROSA CANEDO

TESE DE DOUTORADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR.

APROVADA POR:

ALEXANDRE RICARDO SOARES ROMARIZ, Dr., ENE/UNB
(ORIENTADOR)

ADOLFO BAUCHSPIESS, Dr., ENE/UNB
(EXAMINADOR INTERNO)

RAFAEL TIMÓTEO DE SOUSA JÚNIOR, Dr., ENE/UnB
(EXAMINADOR INTERNO)

IWENS GERVASIO SENE JÚNIOR, Dr., UFG
(EXAMINADOR EXTERNO)

Brasília, 25 de junho de 2019.

FICHA CATALOGRÁFICA

CANÊDO, DANIEL ROSA
SISTEMA DE DETECÇÃO E CLASSIFICAÇÃO DE INTRUSÃO EM REDES AD HOC UTILIZANDO REDES NEURAIS ARTIFICIAIS E ALGORITMO K-MÉDIAS [Distrito Federal] 2019.
xvi, 143 p., 210 x 297 mm (ENE/FT/UnB, Doutor, Engenharia Elétrica, 2019).
Tese de Doutorado em Engenharia de Sistemas Eletrônicos e Automação - Universidade de Brasília, Faculdade de Tecnologia.
Departamento de Engenharia Elétrica

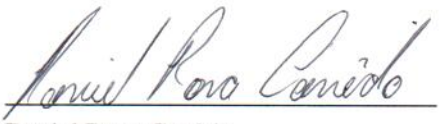
1. Redes Wireless Ad Hoc	2. Multilayer Perceptron
3. K-Médias	4. Sistema de Detecção de Intrusão
I. ENE/FT/UnB	II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

CANÊDO, D.R. (2019). *SISTEMA DE DETECÇÃO E CLASSIFICAÇÃO DE INTRUSÃO EM REDES AD HOC UTILIZANDO REDES NEURAIS ARTIFICIAIS E ALGORITMO K-MÉDIAS*. Tese de Doutorado em Engenharia de Sistemas Eletrônicos e Automação, Publicação PPGEA.TD-141/2019, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 143 p.

CESSÃO DE DIREITOS

AUTOR: Daniel Rosa Canêdo
TÍTULO: SISTEMA DE DETECÇÃO E CLASSIFICAÇÃO DE INTRUSÃO EM REDES AD HOC UTILIZANDO REDES NEURAIS ARTIFICIAIS E ALGORITMO K-MÉDIAS.
GRAU: Doutor ANO: 2019



Daniel Rosa Canêdo
Depto. de Engenharia Elétrica (ENE) - FT
Universidade de Brasília (UnB)
Campus Darcy Ribeiro
CEP 70919-970 - Brasília - DF - Brasil

Dedicatória

Dedico este trabalho primeiramente a Deus, senhor do universo, e em especial a minha esposa, Karla Cristina Duarte Borba e minhas filhas, Lauana Duarte Canedo e Nicole Duarte Canedo.

DANIEL ROSA CANÊDO

Agradecimentos

Agradeço a Deus, o Grande Arquiteto do Universo.

Ao meu orientador, Prof. Dr. Alexandre Ricardo Soares Romariz, por acreditar na minha capacidade.

Aos demais professores que constituíram a banca examinadora para avaliar esta tese de doutorado.

Aos colegas do Instituto Federal de Goiás - Câmpus Luziânia – IFG.

A minha família, pelo incentivo, apoio e torcida pelo êxito desse trabalho.

Um agradecimento especial a minha esposa, Karla Cristina Duarte Borba e minhas filhas, Lauana Duarte Canedo e Nicole Duarte Canedo, por estarem sempre ao meu lado e serem a principal fonte de inspiração na construção deste trabalho.

A todos que contribuíram de forma direta e indireta para realização desse trabalho.

DANIEL ROSA CANÊDO

RESUMO

O acelerado desenvolvimento tecnológico na infraestrutura de tecnologias móveis. O aumento no uso de redes locais sem fio e o uso de serviços de satélites também são perceptíveis. A alta taxa de utilização de dispositivos móveis para diversos fins traz a necessidade de monitorar as redes sem fio. Com esta quantidade de informações transmitidas em redes sem fio se faz necessário identificar de forma rápida e eficiente o tráfego normal e anormal dessas redes, para que seus administradores possam agir. Esta tese apresenta a proposta de um Sistema de Detecção e Classificação de Intrusão em Redes Sem Fio Ad Hoc local composto de duas etapas, baseado em agrupamento de dados através do algoritmo K-Médias e pela Rede Neural Artificial *Multilayer Perceptron*, para a detecção e classificação de anomalias causadas por ataques a estas redes. Estas estratégias são baseadas em algoritmos inteligentes, que são capazes de minimizar as dificuldades que administradores possuem em controlar os diversos integrantes destas redes, bem como na identificação de diversas anomalias. Os algoritmos presentes nesta proposta representam técnicas de classificação, as quais possuem a característica de aprendizagem não supervisionada e supervisionada. O sistema proposto organiza os dados da Rede Ad Hoc em 25 *clusters* em 14,82 segundos através da utilização do K-Médias e possui taxa de classificação de 98,07% utilizando a Rede Neural *Multilayer Perceptron*, tornando-se viável para o processo de classificação de anomalias em Redes Sem Fio Ad Hoc.

ABSTRACT

The accelerated technological development in the infrastructure of mobile technologies. The increase in the use of wireless local area networks and the use of satellite services are also noticeable. The widespread use of mobile devices for various purposes brings the need to monitor wireless networks. With this amount of information transmitted over wireless networks it is necessary to quickly and efficiently identify the normal and abnormal traffic of these networks so that their administrators can act. This thesis presents the proposal of an Intrusion Detection and Classification System in Local Wireless Ad Hoc Networks composed of two stages, based on data grouping through the K-Means algorithm and the Multilayer Perceptron Artificial Neural Network, for the detection and classification of anomalies caused by attacks on these networks. These strategies are based on intelligent algorithms, which are able to minimize the difficulties that administrators have in controlling the various members of these networks, as well as in the identification of several anomalies. The algorithms present in this proposal represent classification techniques, which have the characteristic of unsupervised and supervised learning. The proposed system organizes the Ad Hoc Network data in 25 clusters in 14.82 seconds using the K-Means and has a classification rate of 98.07% using the Perceptron Multilayer Neural Network, making it feasible for the Anomaly Classification in Ad Hoc Wireless Networks

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS

Siglas

Anatel	Agência Nacional de Telecomunicações
ABR	*Associativity-Based Routing*
ACK	*Acknowledgement*
AID	*Association Identity*
AODV	*Ad Hoc On-Demand Distance Vector Routing*
API	*Application Programming Interface*
BPSK	*Binary Phase Shift Keying*
BSS	*Basic Service Set*
CCK	*Complementary Code Keying*
CERT.br	Centro de Estudos, Respostas e Tratamento de Incidentes de Segurança no Brasil
CGSR	*Clusterhead Gateway Switch Routing*
CSMA/CA	*Carrier Sense Multiple Access / Collision Avoidance*
CTS	*Clear to Send*
DCF	*Distributed Coordination Function*
DIFS	*Distributed Coordination Function Interframe Space*
DNS	*Domain Name System*
DoS	*Denial of Service*
DSDV	*Destination-Sequenced Distance-Vector Routing*
DSR	*Dynamic Source Routing*
DSSS	*Direct Sequence Spread Spectrum*
EAP	*Extensible Authentication Protocol*
EMA	Erro Médio Absoluto
FEC	*Forward Error Correction*
FHSS	*Frequency Hopping Spread Spectrum*
FTP	*File Transfer Protocol*
HTTP	*Hypertext Transfer Protocol*
HTTPS	*Hypertext Transfer Protocol Secure*
IBSS	*Independent Basic Service Set*
IEEE	*Institute of Electrical and Eletronics Engineers*
IETF	*Internet Engineering Task Force*
IP	*Internet Protocol*
K-Means	K-Médias
KDD	*Knowledge Discovery and Data*
LMR	*Lightweight Mobile Routing*
LSTM	*Long Short-Term Memory*
LVQ	*Learning Vector Quantization*
MAC	*Media Access Control*
MAO	Mapas Auto-Organizáveis
MLP	*Multilayer Perceptron*
MRQE	Média da Raíz Quadrada do Erro
OFDM	*Orthogonal Frequency-Division Multiplexing*
PAN	*Personal Area Network*
PC	Computadores Pessoais

QPSK	*Quadrature Phase Shift Keying*
RREQ	*Route Request*
RSN	*Robust Security Network*
RTS	*Request to Send*
SDI	Sistema de Detecção de Intrusão
SIFS	*Short Interframe Space*
SMTP	*Simple Mail Transfer Protocol*
SSE	*Sum of Square Erros*
SSR	*Signal Stability Routing*
SVM	*Support Vector Machine*
TCP	*Transmission Control Protocol*
TORA	*Temporally Ordered Routing Algorithm*
UDP	*User Datagram Protocol*
Weka	*Waikato Environment for Knowledge Analysis*
WEP	*Wired Equivalent Provacy*
WMN	*Wireless Mesh Network*
WPA	*Wi-Fi Protected Access*
WRP	*Wireless Routing Protocol*

1 INTRODUÇÃO

Na última década pode-se perceber um grande avanço tecnológico em tecnologias móveis e sua infraestrutura, bem como o aumento da utilização de redes locais sem fio e também da utilização de serviços oriundos de satélites, tanto em ambientes organizacionais, quanto em ambientes residenciais. Isto faz com que informações possam ser criadas, transmitidas e acessadas de forma mais rápida e em qualquer lugar a qualquer momento, bastando apenas ter acesso à infraestrutura de redes móveis. A popularização de dispositivos móveis, principalmente telefones celulares e *tablets*, juntamente com a elevação das telecomunicações, essencialmente de redes celulares, são alguns dos motivos que alavancaram o crescimento da utilização de redes sem fio. Neste tipo de rede não há conexão física entre os dispositivos envolvidos na comunicação, a qual é feita através de ondas eletromagnéticas que trafegam pelo espaço. De acordo com a Anatel(Agência Nacional de Telecomunicações) o Brasil obteve no ano de 2015 257,79 milhões de acessos móveis em operação, sendo que os acessos pré-pagos correspondem a 71,58% do total de acessos, enquanto que os acessos pós-pagos correspondem a 28,42% [11].

Há ainda uma migração do uso de Computadores Pessoais(PC) para dispositivos móveis, como os celulares, *tablets* e *notebooks*. Pesquisa realizada pela empresa IDC Brasil afirma que no último trimestre do ano de 2014 o Brasil obteve 1.637 milhão de computadores, sendo cerca de 600 mil desktops e 1.037 milhão de *notebooks*. Os dados da Anatel, revelam que o Brasil terminou o ano de 2015 com 257,8 milhões de dispositivos móveis, sendo distribuídos entre celulares e *tablets* [11]. As pessoas estão utilizando cada vez mais dispositivos tais como *smartphones* e *tablets* com acesso à Internet. Estes dispositivos pertencem a arquitetura de redes sem fio. Usando essas redes sem fio, os usuários geralmente conseguem obter acesso à Internet com planos mais baratos aos utilizados em redes de telefonia celular. Atualmente estes dispositivos móveis atuam basicamente como computador de pequeno porte, sendo possível realizar todas as ações, entre outras comumente realizadas em um Computador Pessoal, como por exemplo: envio de e-mail através de aplicativos móveis; utilização de um sistema operacional para dispositivos móveis por exemplo *Android e IOS*; visualização de vídeos através de softwares específicos para dispositivos móveis; utilização de serviços de internet como: sistemas web, transações financeiras, compras on-line e outros.

Estes dispositivos móveis fazem parte de redes sem fio, bem como de atuadores sem fio oferecendo tecnologias de comunicação para ferramentas de automação incorporadas à Internet das Coisas em diversos ambientes [12].

A grande taxa de utilização de dispositivos móveis para diversas finalidades explica a importância de se monitorar essa infraestrutura, pois apresenta a transmissão em grande escala de informações, as quais em determinados momentos podem ser restritas. O conjunto deste sistema móvel, determinado tanto pelo *software*, quanto pelo *hardware* utilizado, é relativamente frágil no que se refere à segurança, devido principalmente à característica do seu meio de transmissão,

mas também pelo dinamismo de acesso a este sistema. Portanto tem-se a necessidade de tentar identificar os tráfegos normais e anormais destas redes sem fio para que seus administradores possam tomar ações.

As redes sem fio são classificadas em redes com ou sem infra-estrutura. Em redes infra-estruturadas a comunicação dos dispositivos móveis é realizada com um ou mais equipamentos centralizadores, denominados de pontos de acesso(*access points*), não havendo comunicação direta entre os dispositivos, sempre usando-se um ponto de acesso como intermediário. As redes sem infra-estrutura, denominadas redes *ad hoc*, são formadas por dispositivos que formam uma rede de forma cooperativa, sendo capazes de estabelecer uma comunicação direta com os dispositivos que estiverem ao seu alcance. Nestas redes não existe o gerenciamento centralizado e cada dispositivo possui a funcionalidade de estação e roteador.

Redes Ad Hoc são definidas como redes de computadores sem fio sem a presença de um componente concentrador, tornando cada nó da Rede responsável pelo roteamento e controle de acesso ao meio e gerenciando algumas características da rede como: baixa taxa de transmissão; probabilidade de erro; variações no meio de transmissão. Essas redes são formadas em ambientes onde há necessidade de comunicação, mas há uma inoperabilidade de redes sem fio com estrutura, tornando as redes Ad Hoc de natureza temporária e complexa [13].

Com o aumento da interconexão entre redes, estruturadas e sem fio, a segurança da informação tornou-se uma necessidade. As redes estão sujeitas a vários tipos de ataques que podem ter origem interna ou externa, alguns com objetivos de paralisar serviços, outros com a intenção de roubar informações e em outros casos, apenas por diversão dos atacantes. Além disso, até pouco tempo, as redes eram restritas a computadores, agora aceitam vários tipos de equipamentos: sensores, telefones inteligentes, celulares, entre outros. Portanto, as propostas de melhoria de segurança devem considerar a evolução tecnológica que está ocorrendo.

De acordo com o Centro de Estudos, Respostas e Tratamento de Incidentes de Segurança no Brasil(CERT.br), um incidente de segurança é definido como determinado evento adverso, confirmado ou que esteja sob suspeita, relacionado à segurança de sistemas de computação ou de redes de computadores, tanto com infraestrutura ou sem fio [1]. De maneira geral, qualquer situação onde uma informação esteja sob risco é considerada um incidente de segurança, tais como:

- Acesso não autorizado a sistemas de computação;
- *Denial of service*;
- Vírus e códigos maliciosos;
- Uso impróprio de serviços de tecnologia da informação.

No Brasil, o CERT.br registrou 833.775 casos de incidentes de segurança em 2017. Isso representa um aumento de 28% em relação ao ano anterior, sendo registrado 647.112 incidentes

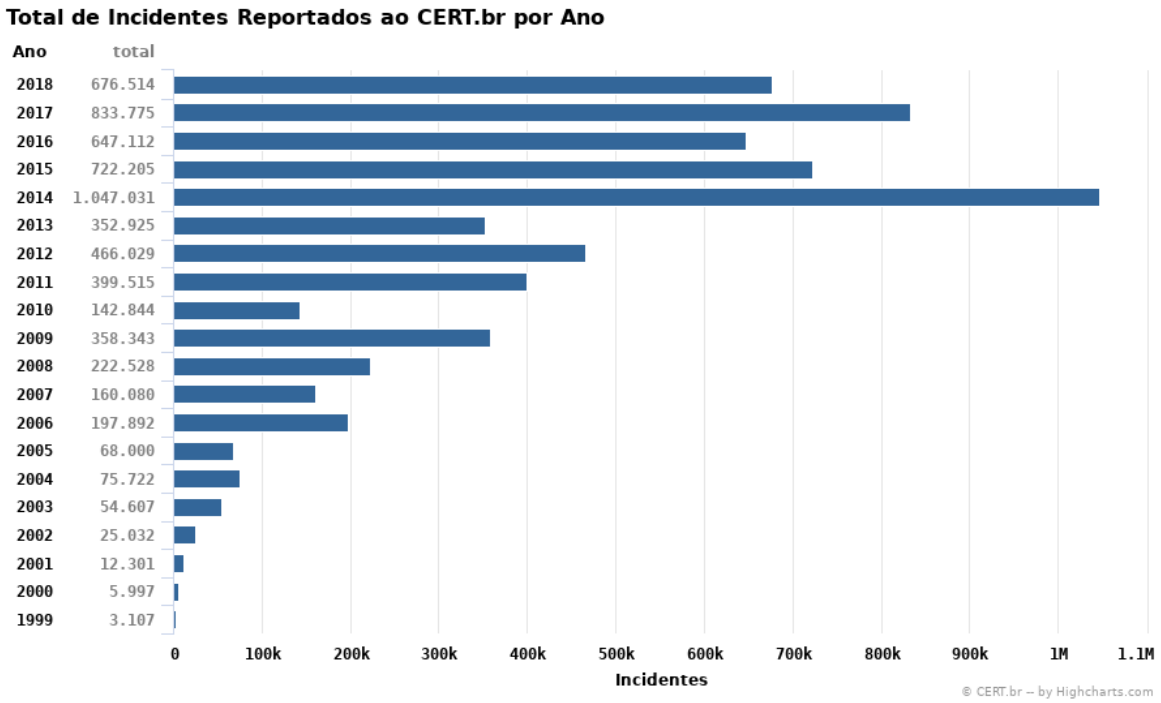


Figura 1.1: Incidentes Reportados ao CERT.br [1].

[1]. A Figura 1.1 mostra que, de 1999 a 2014, o número de incidentes registrados aumentou significativamente, apesar das quedas em alguns anos. De acordo com este cenário, tem-se tornando fundamental a identificação e classificação destes incidentes em redes de computadores, seja estruturada ou sem fio para a definição de políticas de prevenção e mitigação das mesmas.

As redes sem fio Ad Hoc possuem a aptidão de cooperação entre os seu componentes, possibilitando no entanto a existência de diversas vulnerabilidades, sendo passível de ameaças. As ameaças às redes sem fio Ad Hoc podem ser de natureza passiva ou ativa. As ameaças passivas se caracterizam pela espionagem, enquanto que as ameaças ativas alteram ou interrompem o tráfego de mensagens entre os componentes da rede Ad Hoc de forma a comprometer o funcionamento de toda a rede.

De maneira diferente das redes estruturadas, em que os atacantes iniciam o processo de anomalia obtendo acesso físico da rede, ou passando por diversas linhas de defesa, que podem ser: Filtros de roteadores; Regras de agrupamento *broadcast*; *Proxy*; *Firewall*.

Nas redes sem fio Ad Hoc os atacantes realizam o processo de anomalia lançando o ataque contra componentes da rede independente do direcionamento das mensagens. Isto ocorre nas redes sem fio Ad Hoc devido à ausência de um ponto centralizador de distribuição do tráfego ou *gateway*.

As vulnerabilidades observadas em redes estruturadas são as mesmas expostas às redes sem fio Ad Hoc, sendo possível a aplicação de ataques como *spoofing*, *replay*, *denial of service*, etc. As vulnerabilidades em redes sem fio Ad Hoc são classificadas em três grupos, em relação ao seu contexto, ao processo de roteamento e autoconfiguração e ao processo de funcionamento dos

componentes da rede [14].

Uma intrusão é definida como a realização de determinadas ações cujo objetivo é comprometer as propriedades de confidencialidade, integridade e disponibilidade de recursos da rede de computadores, seja de arquitetura estruturada ou sem fio. Um Sistema de Detecção de Intrusão - SDI - deve ser capaz de identificar ações maléficas, no entanto não deve comprometer o funcionamento destas redes. O Sistema de Detecção de Intrusão por sua vez deve consumir poucos recursos computacionais, os quais levarão a um prejuízo de utilização de serviços pelos usuários, como também da utilização em grande escala da largura de banda da Rede [14].

A confidencialidade, integridade e disponibilidade de recursos representam fatores vitais para a segurança da informação. Uma ação maléfica ou não intencional pode comprometer o sistema, caracterizando uma intrusão. O Sistema de Detecção de Intrusão deve conseguir identificar essa ação, mas sem comprometer o funcionamento normal da rede. Um Sistema de Detecção de Intrusão é uma ferramenta de segurança que, como outras medidas, a exemplo de filtros de roteadores, regras de agrupamentos *broadcast*, *proxy*, *firewalls*, destinam-se a reforçar a segurança da informação em sistemas de comunicação [15].

Dependendo da abordagem usada para detectar atividades suspeitas, o Sistema de Detecção de Intrusão pode ser classificado em duas categorias: detecção baseada em anomalia e detecção baseada em assinatura. O primeiro monitora as atividades na rede para detectar desvios efetivos de um comportamento considerado normal. O último consiste em procurar perfis de ataque conhecidos.

Comparando essas duas categorias, pode-se dizer que uma desvantagem da abordagem baseada em anomalias é o alto número de alarmes falsos positivos, e que a base de assinatura exige o conhecimento prévio dos perfis de ataques. No que diz respeito às vantagens, a primeira abordagem é capaz de detectar ataques desconhecidos, enquanto que o segundo é um método de baixa intensidade de computação.

Os Sistemas de Detecção de Intrusão são usados para monitorar, avaliar e informar violações de segurança que podem ser intencionais ou não. No entanto, as técnicas de detecção e prevenção não avançam no mesmo ritmo dos processo de intrusão, devido a diferença de especialistas para evoluírem propostas de intrusão em relação à evolução de técnicas de segurança como por exemplo o Sistema de Detecção de Intrusão.

A análise do tráfego de rede nas redes Ad Hoc é dificultada pela falta de gerenciamento central. Outra característica importante a considerar é a alta mobilidade dos componentes da rede Ad Hoc, já que pode-se entrar e sair da rede sem restrições. Outro aspecto é que os componentes da rede Ad Hoc são na maioria das vezes dispositivos móveis, que possuem restrições em seu estado ativo, pois dependem da energia de seus recursos. Estas propriedades das redes Ad Hoc remetem que os Sistemas de Detecção de Intrusão tradicionais não são usados diretamente e também tem influência direta no funcionamento da rede. Atualmente existem diversas propostas envolvendo Sistemas de Detecção de Intrusão em redes Ad Hoc ou métodos de classificação de tráfego oriundo de redes Ad Hoc [16, 17, 18, 19, 15, 20, 21].

Nesta tese é apresentada uma proposta de um Sistema de Detecção e Classificação de Intrusão em redes Ad Hoc. Este sistema proposto faz uso de técnicas de inteligência computacional com aprendizagem supervisionada e não supervisionada, que são respectivamente Redes Neurais Artificiais do tipo *MultiLayer Perceptron* e algoritmo K-Médias. Este Sistema de Detecção e Classificação é executado de forma local, ou seja, caracteriza-se pela execução local em cada dispositivo pertencente a rede Ad Hoc com o objetivo de não trazer prejuízos aos mesmos. Este sistema, no entanto, preocupa-se em informar ao administrador da rede Ad Hoc, bem como, também ao administrador do dispositivo a respeito dos quadros da rede classificados como anômalos para que possa tomar decisões afim de tornar a rede Ad Hoc estável. Àqueles quadros que não são classificados são armazenados para futuras análises quando possível.

O Sistema de Detecção e Classificação de Intrusão proposto é caracterizado por duas etapas, sendo a primeira responsável pela organização dos dados analisados em grupos, através da similaridade encontrada entre os dados utilizando o algoritmo *K-Médias*. A segunda etapa é responsável pela classificação ou reconhecimento de anomalias através da aplicação de Redes Neurais Artificiais com o algoritmo *Multi Layer Perceptron* levando em consideração informações de grupo determinado na etapa anterior. Por fim, em caso de identificação de anomalias no tráfego da rede sem fio Ad Hoc, o Sistema realizará a ação de incluir informações de determinado quadro na sua estrutura de *log* para posterior análise dos administradores da Rede. Os dados considerados normais poderão compor posteriormente a base de treinamento da Rede Neural, assim como os dados classificados corretamente.

1.1 OBJETIVOS

Esta tese tem como objetivo uma proposta de metodologia de uma solução para detecção e classificação de intrusão em redes Ad Hoc, utilizando estratégias de inteligência computacional capazes de realizar a classificação de anomalias nos componentes de redes sem fio Ad Hoc.

Os objetivos específicos são:

- Avaliar de maneira individual técnicas de inteligência computacional em relação às bases de dados escolhidas;
- Elaborar o Sistema de Detecção e Classificação de redes Ad Hoc;
- Avaliar o sistema para dados coletados de redes sem fio Ad Hoc, bem como para base de dados de redes de computadores comumente utilizada para análise de Sistemas de Detecção de Intrusão.

1.2 JUSTIFICATIVA

O Sistema de Detecção e Classificação de Intrusão proposto atua no ambiente de redes sem fio Ad Hoc utilizando técnicas de inteligência computacional supervisionadas e não supervisionadas. Este sistema, no entanto, se compromete a atingir seu principal objetivo que é obter uma taxa de classificação dos quadros da rede viável, sem prejuízos à rede Ad Hoc. Assim, para conquistar a qualidade do sistema é necessário um estudo aprofundado de propostas de Sistema de Detecção de Intrusão que atuam no ambiente de redes Ad Hoc, bem como, da avaliação do mesmo utilizando dados reais oriundos de implementações de redes Ad Hoc.

As redes sem fio podem ser encontradas na maioria dos ambientes do dia a dia das pessoas, desde em residencias até em ambientes públicos como aeroportos, shoppings e etc. Através das redes sem fio e com a utilização dos dispositivos móveis pode-se ter acesso a Internet, assim como os serviços providos pela mesma tais como: gerenciador de e-mail; acesso a recursos de vídeo e aúdio; acesso a sistemas web; transações financeiras; compras on line; acesso a mídias digitais. Um dos motivos que justifica o desenvolvimento deste trabalho é o fato do crescimento da utilização das redes sem fio, observando sua fragilidade em relação à segurança devido à propriedade do seu meio de transmissão, bem como, a possibilidade de alta mobilidade de seus dispositivos. Estas características dificultam tarefas de prevenção e detecção de incidentes a serem realizadas pelos administradores destas redes.

As redes Ad Hoc, entretanto, são tipos de redes sem fio que têm como principal característica a ausência de componente concentrador, sendo que cada dispositivo da rede tem a funcionalidade de roteamento e de gerenciamento da mesma. Estas redes, no entanto, são mais baratas e fáceis de serem implementadas, sendo utilizadas preferencialmente em ambientes em que se tem uma inoperabilidade de redes infraestruturadas, tornando-as temporárias e complexas. Assim, percebe-se que as redes sem fio Ad Hoc asseguram a existência de diversas vulnerabilidades, sendo passível de ameaças de natureza passiva e/ou ativas comprometendo o funcionamento da rede e de seus dispositivos. Percebe-se um crescimento da utilização de redes Ad Hoc não somente em ambientes complexos, mas também em ambientes corporativos, em eventos e também em processos de automação presentes em Internet das Coisas para diversos ambientes. Este crescimento contribui para a possibilidade do surgimento de incidentes, sendo necessário a utilização de ferramentas capazes de identifica-los em tempo hábil para que administradores possam tomar decisões de prevenção e mitigação na rede.

Sistemas de Detecção de Intrusão são ferramentas que contribuem para garantir a segurança nas redes de computadores, sendo que sua implementação é fundamentada na política de segurança do ambiente com o objetivo de manter ativos os serviços disponibilizados pelas redes de computadores, sem prejuízo às redes. Sistemas de Detecção de Intrusão em redes Ad Hoc são caracterizados por possuirem baixa intensidade computacional, sem prejudicar a própria rede e seus dispositivos e com baixo consumo energético dos dispositivos, tornando inviável a utilização de Sistemas de Detecção de Intrusão tradicionais nestas redes.

A maioria das propostas de sistema de detecção e classificação de intrusão em redes Ad Hoc ou de métodos de classificação [16, 17, 18, 19, 15, 20, 21] têm como principal obstáculo a durabilidade da energia do recurso computacional, sendo então frequentemente utilizadas nestas propostas técnicas de inteligência computacional capazes de analisar, aprender e identificar anomalias. No entanto propostas de sistemas de detecção e classificação de intrusão em etapas estão cada vez mais se preocupando inicialmente em ter um melhor aproveitamento do recurso computacional móvel, através da utilização de técnicas com baixo consumo de processamento, que é responsável pela identificação de uma possível situação de risco. Ficando as demais etapas responsáveis pela classificação das possíveis anomalias. O resultado destes sistemas propostos contribui com os administradores de redes de computadores na escolha de políticas de segurança com o objetivo de prevenir e/ou minimizar os danos causados pelas anomalias.

1.3 TRABALHOS RELACIONADOS

Na literatura pode-se encontrar alguns trabalhos de classificação de tráfego de redes sem fio, os quais podem ser aplicados em Sistemas de Detecção de Intrusão. Essas propostas fazem uso de métodos de aprendizado supervisionados e não-supervisionados.

A proposta de Chandrashekar [16] realiza uma abordagem geral dos diversos métodos de classificação, utilizando dados de alta dimensão e de uma técnica de seleção de variáveis com o objetivo de redução de tempo de computação e melhorar a taxa de aprendizagem.

A proposta de Bhattacharya [17] utiliza uma combinação de métodos de seleção para classificação de anomalias de *Denial Of Service* em redes de computadores, mostrando a eficiência do processo de seleção de recursos para detecção de DoS.

Vo [22] aplica técnicas de aprendizado de máquina supervisionadas e não supervisionadas para prever a tendência de séries temporais, por meio da utilização do algoritmo K-Médias para agrupar dados com similaridade e máquina de vetores para treinamento treinar e testar os dados.

O trabalho de Vlăduţu [18] também utiliza de métodos de classificação não supervisionado e supervisionado para classificar uma coleção de dados de pacotes oriundos da Internet.

Em [19], Nishani apresenta os modelos mais relevantes para a construção de Sistemas de Detecção de Intrusão, incorporando aprendizado de máquina no cenário de redes sem fio Ad Hoc. Os métodos de aprendizagem de máquinas realizam abordagem de classificação, de mineração de regras de associação, Redes Neurais Artificiais e aprendizado com base em instâncias.

Gogoi apresenta a proposta [15] de um método de detecção de intrusão híbrido em vários níveis que usa uma combinação de métodos supervisionados, não supervisionados e baseados em dados discrepantes para melhorar a eficiência da detecção de ataques novos e antigos.

Govindarajan apresenta uma proposta [20] de dois métodos de classificação envolvendo *Perceptron Multicamada* e função de base radial. Propõe-se neste trabalho uma arquitetura híbrida

envolvendo ambos os classificadores para sistemas de detecção de intrusão.

EdWilson [21] propõe um Sistema de Detecção de Intrusão híbrido, em que realiza-se um processamento de sinais através da utilização de transformações *Wavelets* e posteriormente a classificação das anomalias utilizando Redes Neurais Artificais.

EdWilson [8] propõe a elaboração de uma base de dados reais de tráfego de redes sem fio, a qual será utilizada na avaliação de Sistemas de Detecção de Intrusão - SDI. Estes dados por sua vez sofrem um pré-processamento para posteriormente serem classificados por técnicas de reconhecimento de padrões, como por exemplo Redes Neurais Artificiais.

A Tabela 1.1 apresenta de maneira objetiva os principais requisitos dos trabalhos relacionados para a classificação de anomalias.

Tabela 1.1: Requisitos de Trabalhos Relacionados

Trabalho Relacionado	**Técnica de Inteligência Computacional**	**Etapas**	**Base de Dados**
Chandrashekar [16]	Função de Base Radial(RBF), Máquina de Vetores de Suporte(SVM), Redes Neurais	1	Dados de expressão gênica
Bhattacharya [17]	Máquina de Vetores de Suporte(SVM)	3	KDD 99
Vo [22]	K-Médias, Máquina de Vetores de Suporte(SVM)	2	Dados de séries temporais financeiros obtidos do Yahoo Finance
Vlăduţu [18]	K-Médias, J48	2	Tráfego de rede com acesso a internet
Gogoi [15]	MLH	3	KDD 99
Govindarajan [20]	*Perceptron Multicamada(MLP)* e função de base radial(RBF)	2	Sistema imunológico desenvolvido na Universidade do Novo México
EdWilson [21]	*Perceptron Multicamada(MLP)*	2	KDD 99
Nossa Proposta	*Perceptron Multicamada(MLP)* e Algoritmo K-Médias	2	KDD 99 e Redes *wireless* e Ad Hoc

É esperado que a maior parte do tempo, uma rede Ad Hoc funcione em condição normal, que os ataques aconteçam durante pequenos períodos de tempo. É interessante então haver um mecanismo que indique condições anômalas na rede, mas que demande baixo poder computacional. Por isso, a abordagem proposta com o uso do algoritmo K-Médias e Redes Neurais Artificiais *MultiLayer Perceptron* é a principal contribuição desta tese.

1.4 TRABALHOS PUBLICADOS

A seguir são apresentadas as publicações resultantes deste trabalho:

Canêdo, D. R., Romariz, A. R. S. *Sistemas Inteligentes em Sistemas de Detecção de Intrusão para Redes Ad Hoc.* III WPGEA Workshop do Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos e de Automação, UnB, Brasília, 2014.

Canêdo, D. R., Romariz, A. R. S. *Análise de Dados de Redes Sem Fio Utilizando Inteligência Computacional.* IV WPGEA Workshop do Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos e de Automação, UnB, Brasília, 2015.

Canêdo, D. R., Romariz, A. R. S. *Análise de Dados de Redes Sem Fio Utilizando Algoritmos de Classificação.* V WPGEA Workshop do Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos e de Automação, UnB, Brasília, 2016.

Canêdo, D. R., Romariz, A. R. S. *Data Analysis of Wireless Networks Using Computational Intelligence.* Digital Communication Technology and Network Security - ICINC 2018, Barcelona, 2018.

Daniel R. Canêdo and Alexandre R. S. Romariz, *"Data Analysis of Wireless Networks Using Computational Intelligence,"* Journal of Communications, vol. 13, no. 11, pp. 618-626, 2018. Doi: 10.12720/jcm.13.11.618-626

Canêdo, D. R., Romariz, A. R. S. *Análise de Dados de Redes Sem Fio Utilizando Inteligência Computacional.* VII WPGEA Workshop do Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos e de Automação, UnB, Brasília, 2018.

Canêdo, D. R., Romariz, A. R. S. *Data Analysis of Wireless Networks Using Classification Techniques.* 9th International Conference on Computer Science, Engineering and Applications - CCSEA 2019, Toronto - Canada, 2019.

O artigo Intrusion Detection System in Ad Hoc Networks with Neural Networks Artificial and K-Means Algorithm foi submetido para a Revista IEEE América Latina (ISSN 1548-0992) e está em avaliação.

1.5 CONTRIBUIÇÕES

A contribuição desta tese é apresentar um Sistema de Detecção e Classificação de Intrusão para redes sem fio Ad Hoc. Este sistema por sua vez contribui com a utilização de uma abordagem combinando estratégias de técnicas de inteligência computacional com aprendizagem supervisionada e não supervisionada. Essa proposta é aplicada em cada componente da rede sem fio Ad Hoc, sendo possível realizar o agrupamento do tráfego da rede sem o uso de algum evento externo, ou seja, utiliza-se os dados de forma fiel sem restrições. Após este agrupamento, um segundo método é utilizado para classificar as anomalias, caso existam, exigindo um pouco mais

de recurso computacional. Outra contribuição importante neste trabalho é a possibilidade de utilização deste Sistema de Detecção e Classificação de Intrusão juntamente com outras técnicas de segurança em ambientes de redes Ad Hoc. A avaliação do impacto de gerenciamento de energia para as técnicas de inteligência computacional pesquisadas, bem como a comparação entre elas também são contribuições deste trabalho.

Algumas contribuições secundárias podem ser citadas, tais como revisão bibliográfica contendo o estado da arte sobre redes sem fio Ad Hoc; revisão bibliográfica sobre algumas técnicas de inteligência computacional e Sistemas de Detecção de Intrusão; comparação entre algumas propostas de Sistemas de Detecção de Intrusão em redes Ad Hoc.

1.6 ESTRUTURA DA TESE

Esta tese esta organizada da seguinte forma:

- O Capítulo 2 traz a revisão bibliográfica de redes sem fio Ad Hoc. Neste capítulo também faz-se o estado da arte das principais anomalias existentes em redes sem fio Ad Hoc, bem como as principais estratégias de segurança para as mesmas;
- O Capítulo 3 apresenta a revisão bibliográfica sobre Inteligência Computacional e das seguintes técnicas: Redes Neurais Artificiais e Algoritmo K-Médias;
- O Capítulo 4 trata-se da revisão bibliográfica de Sistemas de Detecção de Intrusão;
- O Capítulo 5 expõe a proposta de implementação de um Sistema de Detecção e Classificação de Intrusão baseado em duas etapas, sendo que a primeira etapa faz uso do algoritmo K-Médias e a segunda de Redes Neurais Artificiais. Esta proposta é a principal contribuição desta tese. Também neste Capítulo apresenta-se os resultados obtidos e a análise dos mesmos, bem como uma comparação com outras propostas e outros método de classificação;
- O Capítulo 6, por sua vez exibe as principais conclusões do trabalho, bem como sugestões de trabalhos futuros.

2 REDES SEM FIO AD HOC

A comunicação de dados é parte fundamental de um sistema de computação. As redes de computadores reúnem dados sobre os mais variados assuntos, desde condições atmosféricas a jogos de computadores. Com o desenvolvimento das fibras ópticas, houve grande aumento da banda disponível para transporte de dados. Isso contribuiu para que as redes pudessem ser conectadas entre si, com apoio à criação de uma grande rede com abrangência mundial[23]. As redes podem ser entidades autônomas, sem dependência de outras redes, com disponibilidade de serviços de comunicação para um determinado grupo de usuários.

A área de redes de computadores é uma das quais mais se desenvolvem tecnologias que visam, primordialmente, prover mecanismos e técnicas para comunicar e integrar as diversas regiões, empresas e cidadãos do mundo. Essas tecnologias utilizam vários meios de transmissão e também proporcionam formas de comunicação rápida, segura e eficiente.

O avanço das tecnologias de informação e comunicação faz com que haja a possibilidade de cenários em que os componentes das redes de computadores tenham um maior dinamismo, possibilitando a construção de redes com mobilidade, sem ter seus componentes presos a uma infraestrutura física. Surgem então as redes de computadores sem fio, possibilitando a construção de redes de computadores com alta mobilidade entre os seus componentes, que podem ser: *notebook*, celulares, tablets e etc. Estas redes por sua vez, são classificadas em redes sem fio com infraestrutura e sem infraestrutura, sendo que o que as difere é a presença de um componente centralizador, em redes sem fio com infraestrutura, capaz de realizar o gerenciamento de transações entre seus componentes.

Neste capítulo é apresentada a fundamentação teórica sobre redes de computadores sem fio Ad Hoc, o protocolo da camada de transporte, o qual faz parte da pesquisa desta tese. Também neste capítulo é feita referência teórica dos requisitos de segurança em redes sem fio Ad Hoc, bem como a descrição das principais anomalias em Redes Ad Hoc.

2.1 PROTOCOLOS TCP/IP

O Departamento de Defesa dos Estados Unidos, no final da década de 60 autoriza a pesquisa e o desenvolvimento de uma rede de computadores com objetivo de interligar os diversos centros de pesquisa. Na oportunidade, com a guerra fria, havia interesse no desenvolvimento de uma infraestrutura com funcionamento independente dos nós da rede, ainda que há possibilidade de alguns centros sofrerem ataques. O protótipo denominado ARPANet é desenvolvido, porém apresentava problemas de estabilidade com constantes quedas. Assim, inicia-se a pesquisa para construir um conjunto de protocolos confiáveis. Este projeto é finalizado na década de 70 com o desenvolvimento dos protocolos TCP/IP[24].

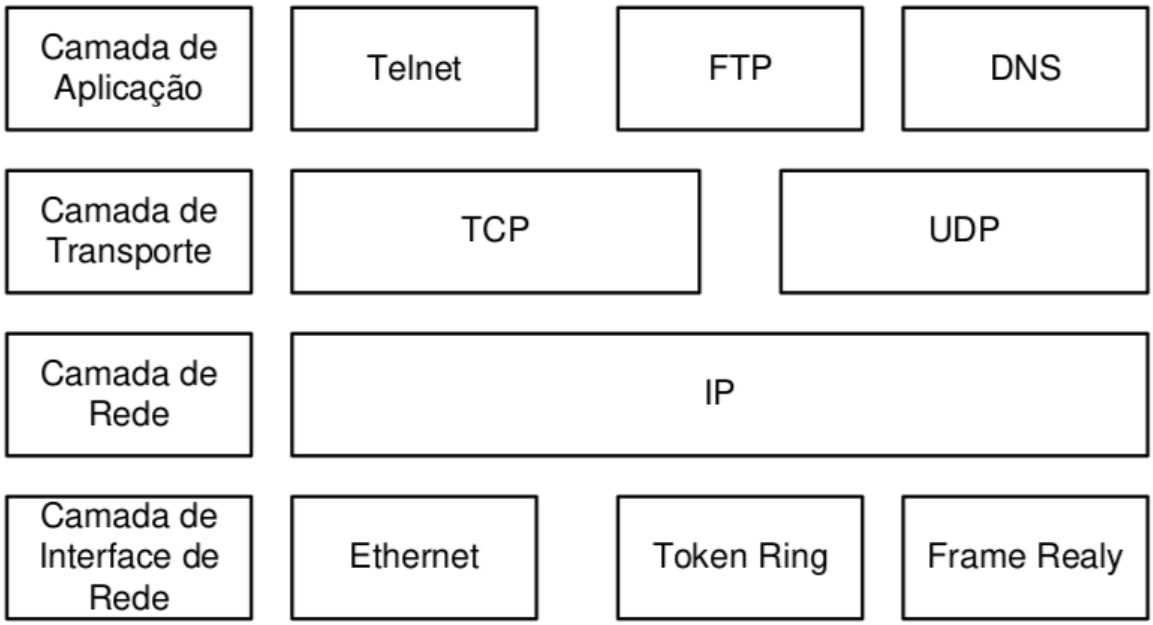


Figura 2.1: Modelo de Referência TCP/IP [2].

O modelo de referência TCP/IP não especifica as camadas mais baixas de rede (física e enlace), o que possibilita diversas redes de computadores, como por exemplo Ethernet, *Frame Relay*, ATM, redes sem fio dentre outras, de realizar o transporte de pacotes TCP/IP. No momento da transmissão de pacotes entre as redes, deve-se garantir que os dados sejam transmitidos, independente do destino. O protocolo IP(*Internet Protocol*) pode entregar os pacotes fora da ordem original em que são criados pelo emissor. A rede pode utilizar diferentes caminhos para entregar os mesmos a um determinado destinatário, fazendo com que as camadas de níveis superiores no destinatário sejam responsáveis por ordenar os pacotes recebidos. O principal objetivo da camada de rede é realizar o roteamento para entrega dos pacotes. A Figura 2.1 apresenta uma visão geral da família destes protocolos, sendo que para este trabalho os dados a serem analisados pertencem a camada de aplicação.

As funções da camada de transporte no modelo TCP/IP são semelhantes às funções apresentadas pelo nível de transporte do RM-OSI. Esta camada é definida pela utilização de dois protocolos: TCP(*Transmission Control Protocol*) e UDP(*User Datagram Protocol*).

O TCP permite a entrega dos pacotes sem erros, através de um fluxo de *bytes* em um canal originado no transmissor para o receptor. É um protocolo orientado à conexão e possui funções de controle de erro, controle de fluxo, sequenciamento e multiplexação do acesso. Antes de realizar a transmissão de dados, recebidos da camada de aplicação, o TCP precisa estabelecer uma conexão entre os dois nós envolvidos na comunicação[25].

O cabeçalho do segmento TCP é constituído pelos campos: números da porta de origem e destino, número de sequência, número de reconhecimento, comprimento do cabeçalho, *flags*, janela de recepção, valor para checagem de dados, identificação para dados urgentes e um campo opcional de opções. Os *flags* são constituídos de seis *bits* assim distribuídos:

- ACK: indica a confirmação de recebimento de um, ou vários, segmentos;
- RST: indica a intenção do emissor em abortar, de forma abrupta, a conexão;
- SYN: utilizado para estabelecer uma conexão;
- FIN: indica a intenção do emissor em finalizar, de forma normal, a conexão;
- PSH: indica que o destinatário deve enviar imediatamente os dados para a camada superior;
- URG: indica que existem dados no segmento que são marcados, pela camada superior, como urgentes.

O estabelecimento de uma conexão TCP inicia-se com a solicitação do cliente, através de um segmento especial, com o *flag* SYN ativado, para o servidor. Caso o servidor aceite a conexão, após fazer a alocação de *buffers* e variáveis do protocolo TCP, este por sua vez retorna para o cliente um segmento com o *flag SYN* ativado. Por fim, o cliente confirma para o servidor, com um segmento com o *flag ACK* ativado. Esta explicação não apresenta itens como número de sequência e número de reconhecimento. Este procedimento utilizado pelo TCP é denominado de "aperto de mão" de três vias. Ao término, os nós possuem uma conexão e estão prontos para transmissão de dados oriundos da camada superior.

O protocolo UDP é mais simples, em se comparando com o protocolo TCP. O UDP não é orientado à conexão, ou seja, os datagramas são encaminhados para a camada de rede sem a garantia de entrega. Esse protocolo por sua vez não garante a ordem de chegada dos datagramas. Essas verificações são de responsabilidade dos protocolos das camadas de aplicação. A partir disto o UDP é utilizado por aplicações que necessitam de entrega imediata dos dados, como transmissão de áudio e vídeo.

A camada de aplicação no entanto possui protocolos de alto nível para permitir a utilização pelos usuários. Algumas aplicações utilizam o mesmo nome dos protocolos (HTTP(*Hypertext Transfer Protocol*), HTTPS(*Hyper Text Transfer Protocol Secure*), FTP(*File Transfer Protocol*), DNS(*Domain Name System*) e SMTP(*Simple Mail Transfer Protocol*)) [26].

2.2 PROTOCOLOS DE REDES SEM FIO

As redes de computadores cabeadas restringem a mobilidade de seus usuários, ficando impossível a movimentação de equipamentos, pois estão presos pela infraestrutura física. As redes sem fio entretanto possibilitam a mobilidade para os usuários de dispositivos computacionais móveis, como *notebook*, celulares, *tablets* e etc. As redes sem fio podem ser divididas em dois grupos: com infraestrutura e sem infraestrutura.

Nas redes sem fio com infraestrutura, toda comunicação é realizada através de um ponto concentrador, como acontece com as redes de comunicação celular, enquanto que nas redes sem

fio sem infraestrutura os nós da rede comunicam-se diretamente, sem a presença de um ponto concentrador. Estas redes são chamadas de redes sem fio Ad Hoc. A comunicação pode ainda ser direta entre os nós vizinhos, ou por múltiplos saltos, sendo que nestes casos os nós também funcionam como roteadores na rede.

As redes sem fio Ad Hoc são divididas nos seguintes tipos [14]:

- *Mobile Ad Hoc Networks (MANET)*;
- *Vehicular Ad Hoc Networks (VANETs)*;
- *Wireless Sensor Network (WSN)*;
- *Wireless Mesh Networks (WMN)*.

As redes sem fio abordadas nesta tese são as redes MANETs que são padronizadas de acordo com especificações do IEEE(*Institute of Electrical and Eletronics Engineers*). Os trabalhos da primeira versão da família 802.11, iniciaram-se em 1990, sendo especificada a camada física com três alternativas de transmissão: FHSS(*Frequency-Hopping Spread Spectrum*), DSSS(*Direct Sequence Spread Spectrum*) e infravermelho [27].

A faixa de frequência, alocada para os canais, sobrepõe canais próximos. A Tabela 2.1 [10] especifica o padrão com até quatorze canais de frequência, sendo somente três utilizados simultaneamente sem interferência (canais um, seis e onze). O canal seis superpõe aos canais cinco, quatro, três e dois, além dos canais sete, oito, nove e dez. Na literatura, encontram-se vários trabalhos relacionados à alocação de canais, como mostrado em [10], [28], [29] e [30].

Tabela 2.1: Canais do Padrão IEEE 802.11 [10]

Número do Canal	**Frequência Central de Canal (MHz)**
1	2412
2	2417
3	2422
4	2427
5	2432
6	2437
7	2442
8	2447
9	2452
10	2457
11	2462
12	2467
13	2472
14	2484

O padrão IEEE 802.11b é uma evolução do padrão original IEEE 802.11, também conhecido como *Wifi* e possui suporte a taxa de transmissão de até 11Mbps. A partir da revisão do padrão

IEEE 802.11 surgem os seguintes padrões:

- IEEE 802.11b;
- IEEE 802.11a;
- IEEE 802.11g.

A Tabela 2.2 apresenta as características dos padrões do IEEE, enfatizando suas frequências e taxa máxima de transmissão. Deve-se salientar que fabricantes comercializam equipamentos com capacidade maior de transmissão, entretanto esses valores não são reconhecidos pelo IEEE, logo não seguem a padronização e não existe garantia de interoperabilidade com outras marcas.

Tabela 2.2: Padrões IEEE 802.11

Padrão	**Faixa de Frequência em GHz**	**Transmissão Máxima em Mbps**	**Modulação**
IEEE 802.11	2,4 – 2,483	2	FHSS/DSSS
IEEE 802.11a	5,1 - 5,8	54	OFDM
IEEE 802.11b	2,4 – 2,485	11	DQPSK
IEEE 802.11g	2,4 – 2,485	54	OFDM

Os padrões definidos pelo IEEE especificam a taxa máxima de transferência, sendo que alguns fatores podem reduzir consideravelmente a mesma:

- Obstáculos que degradam o sinal (paredes, campos eletromagnéticos, construções prediais);
- Saturação do espectro (este tipo de acesso é compartilhado, com o aumento de número de usuários, a probabilidade de colisão na transferência de dados aumenta);
- Interferência por outras redes (a existência de outras redes, na mesma faixa de frequência, pode degradar o sinal, até mesmo anulá-lo).

Muitas características são compartilhadas entre os padrões, pois todos utilizam o mesmo protocolo de acesso ao meio, o CSMA/CA(*Carrier Sense Multiple Access / Collision Avoidance*). Também utilizam a mesma estrutura de quadros na camada de enlace e possuem capacidade de reduzir a taxa de transmissão para alcançar distâncias maiores, e igualmente permitem modo de infraestrutura e *ad hoc* [26].

A transmissão de dados, entre os nós, é realizada através de um meio compartilhado. A camada MAC(*Media Access Control*) tem como principal função gerenciar o acesso compartilhado ao meio. O mecanismo de controle da camada MAC suporta dois métodos de acesso, distribuído e centralizado, com a possibilidade de ambos coexistirem. Os métodos de acesso determinam quando uma estação da rede tem permissão para utilizá-la.

A decisão de transmissão pode ser tomada de forma individual pela estação, fazendo com que a coordenação seja distribuída. Isso poderá resultar em transmissões simultâneas, que exigirão a

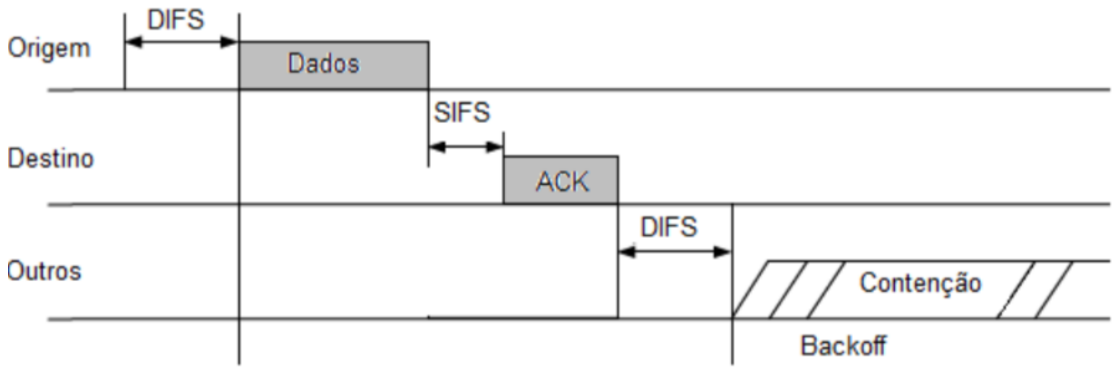


Figura 2.2: Acesso no *Distributed Coordination Function* [2].

retransmissão pelos nós, neste caso, o usuário poderá sofrer uma sensação de lentidão. O modelo centralizado para decisão de transmissão poderá reduzir a ocorrência das colisões. Em ambos os métodos, a estação, quando desejar transmitir na rede, deverá primeiro ouvir o meio e se estiver livre (sem outra estação transmitindo), o nó poderá iniciar o uso da rede.

A função de coordenação distribuída baseia-se no protocolo CSMA/CA para controle do acesso ao meio. As redes sem fio, em modo ad hoc, devem obrigatoriamente utilizar este método. Se o meio estiver ocupado, será necessário aguardar pela duração de DIFS(*Distributed Coordination Function Interframe Space*), que é o tempo entre a transmissão dos quadros. O nó deverá então entrar numa fase de contenção, e tentar acessar o meio novamente após este intervalo aleatório de tempo. Caso o meio continue ocupado, todo este processo será repetido, inclusive a espera aleatória, como mostrado na Figura 2.2 [2].

A estação destino confirma o recebimento de quadros sem erros através de ACK (*Acknowledgement*), após um intervalo de tempo chamado SIFS (*Short Interframe Space*). Se o nó origem não receber o ACK, deduzirá que houve colisão, então iniciará a retransmissão e entrará no processo de retenção. Neste caso, aguardará um tempo aleatório, uniformemente distribuído entre zero e o tamanho da janela de contenção, com objetivo de evitar novas colisões. Apenas a utilização destes procedimentos não resolverá o problema de nós ocultos. Conforme verifica-se na Figura 2.3 [2] este problema acontece pois os nós B e A ouvem um ao outro, assim como os nós B e C. Porém os nós A e C não podem ouvir um ao outro, isto implica que não se dão conta da sua interferência em B. Se o nó A iniciar uma transmissão para B e no mesmo instante o nó C também iniciar uma transmissão para o nó B (pois o nó C não conseguiu identificar que o nó A está utilizando o meio) haverá colisão nas transmissões.

Para melhorar este cenário, desenvolveu-se um mecanismo opcional que envolve a troca de quadros de controle RTS (*Request to Send*) e CTS (*Clear To Send*). Uma estação, quando deseja transmitir na rede, envia um quadro de controle RTS, o qual informa uma estimativa de tempo da futura transmissão. A estação destino, em resposta ao quadro RTS recebido, envia um quadro de controle CTS com a indicação de que está pronta para receber os quadros de dados. Somente após a confirmação pelo nó destino o emissor inicia a transmissão. O quadro RTS possui funcionalidade de reservar o meio para transmissão e verificar se o destinatário está pronto para

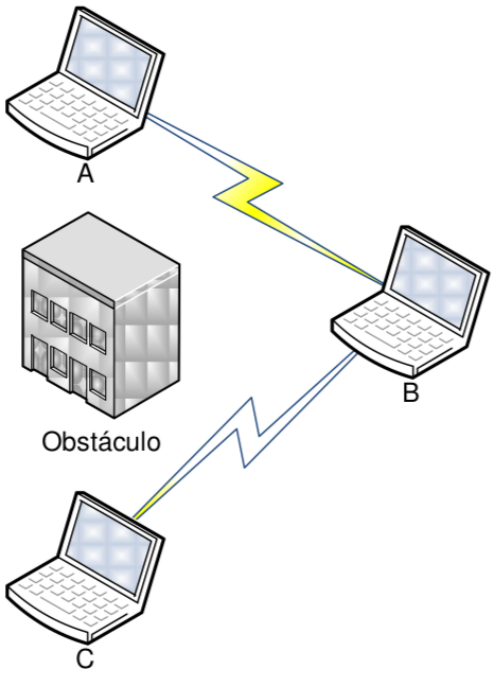


Figura 2.3: Rede Ad Hoc com obstáculo [2].

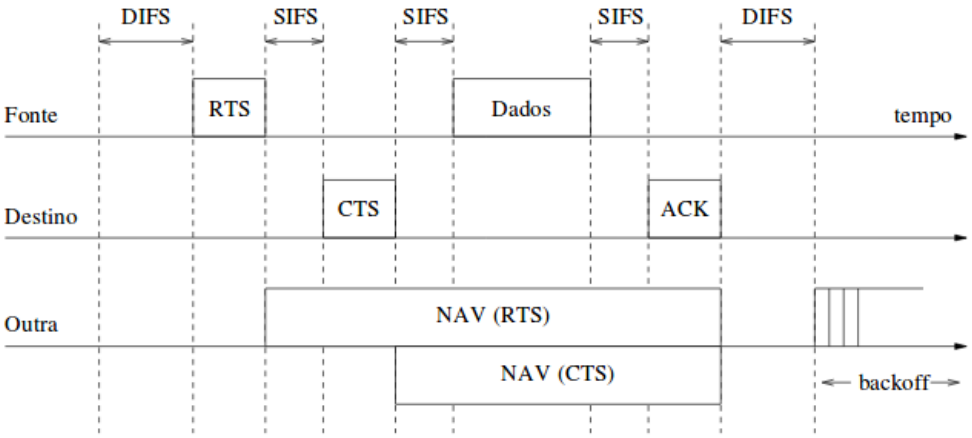


Figura 2.4: Acesso ao Canal com Quadros RTS/CTS [3].

recebimento. Este procedimento é apresentado na Figura 2.4 [3].

O padrão IEEE 802.11 também inclui uma função opcional, chamada Função de Coordenação Centralizada, que diferentemente da DCF(*Distributed Coordination Function*), é um modelo MAC centralizado onde um ponto de acesso elege, de acordo com suas regras, um terminal *wireless* para que este possa transmitir seu pacote.

Conforme a Figura 2.5 [2] o elemento BSS (*Basic Service Set*) é responsável pelo gerenciamento dos nós da rede [31]. Caso um nó da rede encaminhe pacotes para seu vizinho, necessariamente, a transmissão passará e será controlada pelo BSS. A interligação com outros BSS é realizada através do sistema de distribuição (DS). No modo ad hoc, as estações são independen-

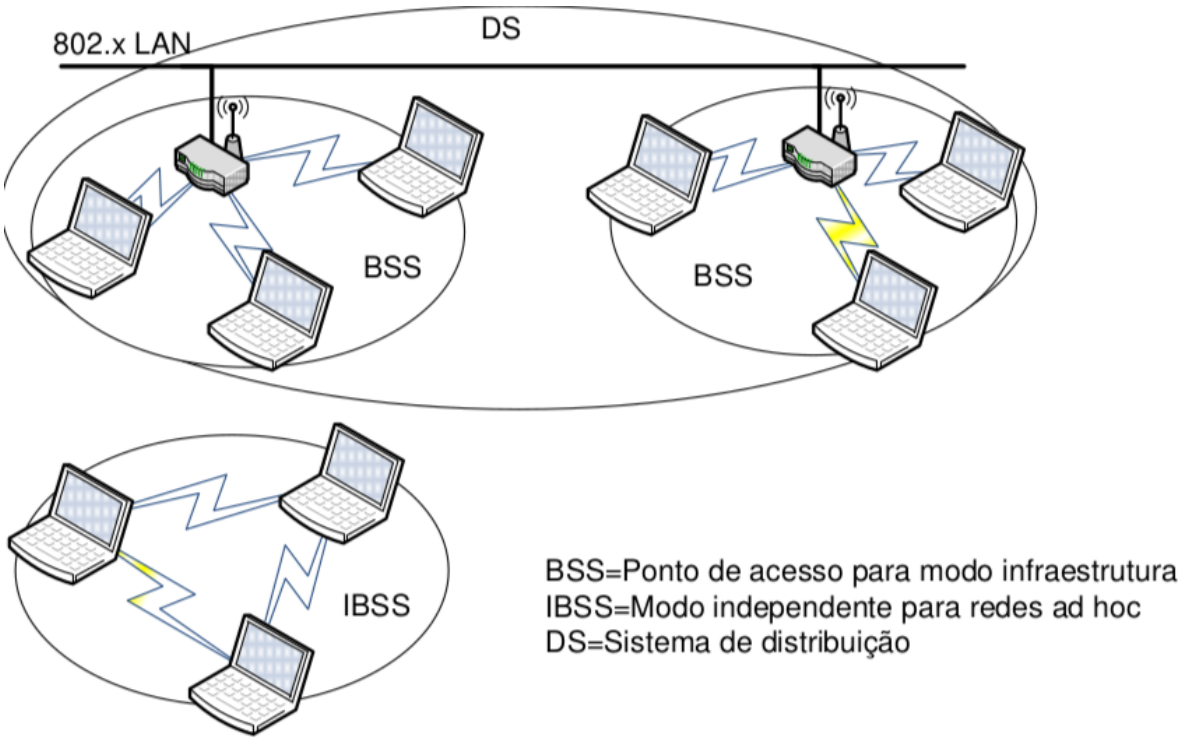


Figura 2.5: Arquitetura Redes Sem Fio [2].

tes e não possuem infraestrutura centralizada, assim, em comunicações entre nós vizinhos, não será necessário um elemento central fazer o encaminhamento de pacotes, desde que ambos os nós estejam acessíveis, e formam então o IBSS (*Independent Basic Service Set*).

O padrão IEEE 802.11a provê melhorias para a camada física, mantendo as outras camadas intactas [32]. A principal vantagem deste padrão é a diminuição de possibilidades de interferência, se comparado com o modelo 802.11. Outro avanço é o aumento da taxa de transferência para o limite teórico de 54Mbps. Os equipamentos podem suportar taxa variável de transmissão (6, 9, 12, 18, 24, 36 e 48Mbps) por fazer uso de técnicas diferentes de modulação. Este padrão utiliza 300MHz de largura de banda, na faixa de 5,4GHz. É variável também a potência máxima que pode ser utilizada. Por exemplo, os canais baixos (5,15 até 5,25GHz) operam com até 50mW, enquanto que os canais intermediários (5,25 até 5,35GHz) operam com potência até 250mW e os canais mais altos (5,725 até 5,825GHz) podem utilizar 1W de potência.

A existência de várias portadoras de baixa velocidade forma um canal de alta velocidade. A modulação OFDM(*Orthogonal Frequency-Division Multiplexing*) define oito canais (sem sobreposição) de 20MHz. Cada um destes canais é dividido em 52 subportadoras (aproximadamente 300KHz cada) que são transmitidas em paralelo. É utilizado o FEC(*Forward Error Correction*), com a inclusão de *bits* adicionais nos dados transmitidos para realizar a detecção de erros, pois o *overhead* produzido não é expressivo se comparado com a banda disponível para transmissão. O OFDM especifica baixa velocidade de transmissão de símbolos, a chance de interferência por propagação de múltiplos percursos também é pequena [32].

A especificação do padrão IEEE 802.11b, da mesma forma que acontece com o IEEE 802.11a, modifica apenas a camada física, quando comparado à especificação inicial. Permite taxas de 1; 2; 5,5 e 11Mbps e possui 14 canais (ou 11 canais em alguns países) com sobreposição. As taxas especificadas no IEEE 802.11b são variantes do IEEE 802.11 que faz uso do CCK(*Complementary Code Keying*), método de modulação baseado em códigos complementares binários. A taxa de 1Mbps é codificada em BPSK(*Binary Phase Shift Keying*), modulação que separa as fases em 180 graus, enquanto que as outras três maiores em QPSK(*Quadrature Phase Shift Keying*), onde são utilizados os parâmetros de fase e quadratura da onda portadora para modular o sinal de informação. O padrão IEEE 802.11g também opera na faixa de 2,4GHz e possui taxa de transmissão de até 54Mbps [32].

2.3 REDES SEM FIO AD HOC

Redes sem fio Ad Hoc, também conhecidas por Manet, são redes sem fio cujos componentes comunicam sem a presença de uma infraestrutura de rede previamente definida. Estas redes por sua vez possuem a característica de serem amplas, pois seus componentes comunicam-se de forma direta, através de uma arquitetura de comunicação ponto-a-ponto. No entanto, pelo aspecto de não possuirem infraestrutura, os serviços de roteamento são definidos de maneira cooperativa, fazendo que cada estação participante da rede sem fio Ad Hoc possa atuar como roteador para outras estações. Desta forma, quando uma estação desejar se comunicar com outra que não esteja ao alcande de seu enlace, esta encaminha seus pacotes para a estação mais próxima do destinatário, que encaminhará os pacotes adiante até serem recebidos pela estação destino. Consequentemente, as redes sem fio Ad Hoc (Manet) são definidas como redes móveis multi-salto, cuja conectividade é realizada pela capacidade de roteamento colaborativo dos componentes das mesmas [14].

A difusão de dispositivos móveis, como telefones celulares, *tablet*, *notebook*, juntamente com o avanço da tecnologia da informação mais precisamente na área de telecomunicações, são algumas das razões para o incentivo de crescimento da utilização das redes sem fio, dentre as quais encontram-se as redes sem fio Ad Hoc. Conforme mencionado nestes tipos de redes não existe a ligação física entre os seus dispositivos sendo feita através de ondas eletromagnéticas que trafegam pelo espaço [33].

As redes Ad Hoc são principalemnte compostas de dispositivos móveis com uma ou mais interfaces de redes sem fio. Normalmente, os enlaces destas redes continuam com uma capacidade notadamente menor que os enlaces em redes com infraestrutura(cabeadas). Esta característica das redes Ad Hoc é percebida não somente pela diferença e limitações em relação a vazão nominal de determinada interface, mas também a fatores próprios destas redes tais como: efeitos de múltiplo acesso, desvanecimento(*fading*), ruídos e interferências presentes em fontes eletromagnéticas exógenas ao sistema, entre outros. Além do mais, os dispositivos móveis são alimentados por fontes de energia portáteis (baterias) que se extinguem com o tempo. Desta maneira, projetos que tem por objetivo potencializar os recursos e serviços das redes Ad Hoc possuem como critérios

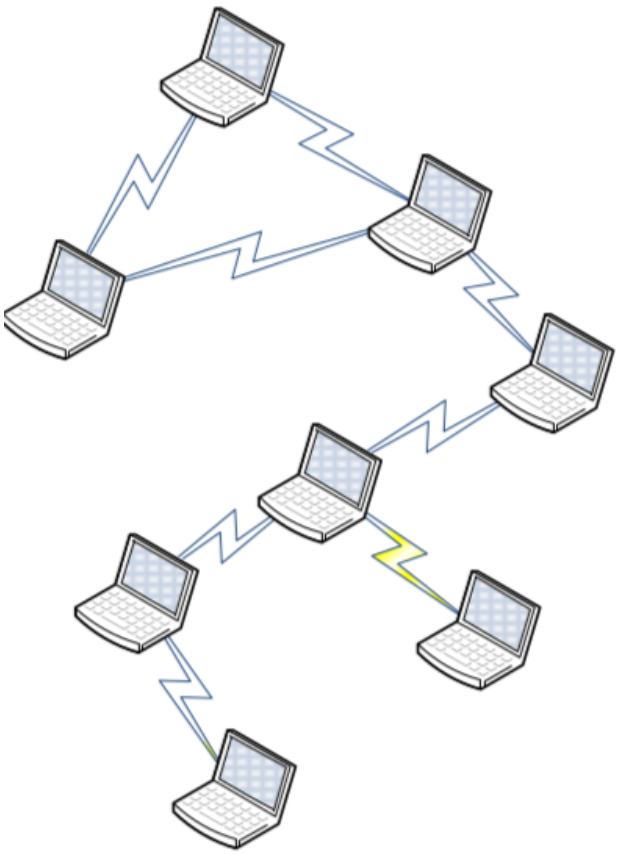


Figura 2.6: Modelo de Rede Ad Hoc [2].

fundamentais o uso eficaz de banda e energia disponíveis [14].

A Figura 2.6 [2] apresenta um modelo de rede sem fio Ad Hoc que permite a comunicação diretamente entre os nós, e estes por sua vez podem realizar o repasse de pacotes através de múltiplos saltos. Cada elemento da rede é responsável pelo encaminhamento de pacotes de seus vizinhos. Cada nó é equipado com uma ou mais interface de rádio, e a cobertura da rede depende diretamente do alcance destes enlaces. Até certo ponto, é possível adicionar mais dispositivos na rede e, consequentemente, aumentar sua cobertura.

Os dispositivos da rede também podem funcionar como roteadores dentro da rede para outros dispositivos, com o encaminhamento de pacotes para o destinatário final. Esta rede pode possuir conexão com rede com infraestrutura(cabeada e sem fio), através de *gateways* [34].

Os dipositivos móveis que formam as redes Ad Hoc podem de maneira constante e a qualquer momento, aparecer, desaparecer e mover-se dentro das mesmas. Deste modo, a aceitação de dispositivos em redes Ad Hoc é construída de maneira dinâmica, tornando a topologia da rede com mudanças contínuas e imprevisíveis. Esta característica da rede Ad Hoc, juntamente com a mobilidade dos dispositivos móveis, associada à confiabilidade e a limitação da banda nos enlaces sem fio, torna a disponibilidade de determinado dispositivo da rede não assegurada. Diante disto, os serviços em uma rede Ad Hoc não podem ser concentrados em um dispositivo centralizador, conforme consta em redes sem fio com infraestrutura. No entanto os serviços de redes Ad Hoc

devem ser dotados de características distribuída e auto-organizada, através da colaboração entre os dispositivos da rede. Essa colaboração faz uso das redundâncias naturais resultantes do modelo de comunicação, proporcionando uma compensação pela ausência de confiabilidade em relação a disponibilidade dos dispositivos individualmente [14].

A habilidade de movimentação dentro da rede Ad Hoc aumenta a inconsistência de algumas informações. A rede Ad Hoc é dinâmica, uma vez que tanto sua topologia quanto seus membros mudam constantemente. No momento em que dispositivos deixam a rede ou passam a fazer parte dela, novas rotas são encontradas para que a comunicação entre os dispositivos seja mantida. Assim, com a mobilidade dos dispositivos, a disponibilidade de serviços não são garantidos [33].

A propriedade dinâmica das redes Ad Hoc identifica-se duas funcionalidades destas redes que são: autoconfiguração e roteamento. O serviço de roteamento refere-se à natureza multi-salto das redes Ad Hoc. Desta maneira, os protocolos de roteamento levam em consideração as constantes mudanças na topologia da rede conforme a mobilidade dos dispositivos. Entretanto o serviço de autoconfiguração relaciona-se com a combinação dos dispositivos na rede, implantando rapidamente e com pouca ou nenhuma intervenção dos usuários [14].

Os dispositivos móveis apresentam o aspecto de portabilidade, que permite que haja limitações de capacidade de armazenamento e de consumo de energia. No entanto, o consumo de energia em ambiente de redes sem fio decorre não somente enquanto o dispositivo está trafegando dados, mas também enquanto este encontra-se inativo, sendo necessárias políticas eficientes de gerenciamento do consumo de energia destes dispositivos [33].

Os meios de comunicação sem fio, já citados anteriormente, possuem como maior desafio o seu gerenciamento de banda. O gerenciamento de banda nos meios de comunicação sem fio têm a funcionalidade de disponibilizar largura de banda baixa, juntamente com a obrigação de compartilhamento desta banda entre os dispositivos da rede. Sabe-se que dispositivos podem frequentemente entrar, sair e mover-se dentro da rede resultando na alta variação da largura de banda disponível. As desconexões ocorrem mais frequentemente que em redes com estrutura(cabeada), pois barreiras naturais, como construções ou acidentes geográficos, podem interferir na transmissão do sinal. Desta forma dispositivos podem deixar a rede a qualquer instante, sendo necessária a restauração de rotas para manutenção da comunicação.

O maior desafio em redes sem fio Ad Hoc poderia ser em relação a segurança [14, 33]. Algumas considerações de destaque para a segurança em redes Ad Hoc são:

- A ausência de entidades centralizadoras;
- As características dos enlaces sem fio;
- A natureza volátil destas redes, as quais podem se dividir em um momento e agrupar-se de maneira imprevisível;
- O fato de não se ter como prever o tamanho de uma rede Ad Hoc.

Há algumas vantagens quanto à utilização de redes sem fio Ad Hoc quando comparadas com as redes com infraestrutura [35]:

- Mobilidade dos dispositivos;
- Economia pela ausência de infraestrutura;
- Menor consumo de energia devido aos enlaces curtos.

As desvantagens no entanto estão na complexidade para os dispositivos computacionais, além do roteamento dos pacotes, devem conter mecanismos de controle de acesso ao meio, autoconfiguração e gerenciar características de redes sem fio tais como [35]:

- Baixa taxa de transmissão;
- Tráfego com probabilidade de erro;
- Grande variação das condições do meio de transmissão.

2.3.1 Aplicações de Redes Sem Fio Ad Hoc

Encontram-se várias demandas atuais e futuras para a tecnologia da informação e comunicação de redes Ad Hoc. A computação móvel encontra-se em evolução, conforme é apresentado em [11], em que observa-se um aumento considerável de utilização de dispositos móveis para uso pessoal e em ambientes corporativos, tendo como consequência a utilização de redes sem fio em grande escala. Redes sem fio começam a requerer tecnologias de rede altamente adaptáveis que permitam o gerenciamento de *clusters* multi-saltos de redes Ad Hoc que operem de maneira autônoma ou conectadas em um ou mais pontos à Internet [14].

Fundamentalmente, o uso de tecnologias de redes Ad Hoc relaciona-se com a formação espontânea de redes. Com certeza, o aspecto de auto-organização faz das redes Ad Hoc uma alternativa flexível para a formação de redes, fazendo com que redes sem fio móveis sejam rapidamente determinadas sem a necessidade de implantação prévia de infraestrutura. Diante disto, há diversos cenários de utilização das redes Ad Hoc em aplicações comerciais, industriais, acadêmicas, eventos, governamentais ou militares, dentre as quais pode-se citar [14]:

- Comunicação inter-grupo e trabalho cooperativo: formação dinâmica de grupos colaborativos de trabalho, em ambientes empresariais, acadêmicos e comerciais, entre outros;
- Redes de área pessoal (*Personal Area Network – PAN*): estabelecimento de comunicação em rede para ambientes de dimensão reduzida através de comunicação máquina-a-máquina, eliminando ou reduzindo a necessidade de instalação de dispositivos para ligação e interligação em rede;

- Intervenções em sítios sem infraestrutura ou cuja infraestrutura tenha sido destruída: aplicação em cenários onde se requeira que o estabelecimento rápido de comunicações através de redes dinâmicas e com sobrevivência, como por exemplo em operações de resgate em sítios de acidentes ou atentados, em incêndios, em desabamentos, em procedimentos de manutenção em sítios remotos, entre outros;
- Redes de sensores: formação de redes entre diversos sensores, que eventualmente se encontram em movimento, para troca e processamento de informações relacionadas com as medidas que estão sendo realizadas;
- Redes em movimento: redes constituídas por sistemas em movimento, tais como aviões, carros em uma estrada ou tropas em um campo de batalha;
- Internet das Coisas: dispositivos móveis, bem como atuadores sem fio oferecendo tecnologias de comunicação para ferramentas de automação incorporadas a Internet das Coisas em diversos ambientes [12]

Complementarmente, outro cenário destaca-se para as tecnologias de informação e comunicação de redes sem fio Ad Hoc, que é a comunicação pervasiva e a formação de redes que tenham acesso ubíquo [36, 37, 38]. Desta maneira, tem-se as redes Ad Hoc com base em malhas que são uma alternativa ou um complemento desenvolvido e barato para as redes móveis celulares [14].

2.3.2 Redes Ad Hoc em Malha

As redes Ad Hoc em Malha, também denominadas de *Wireless Mesh Network - WMN* são redes Ad Hoc com certas particularidades, sendo que seus dispositivos não possuem mobilidade. Geralmente são instalados nos topos de edifícios ou em telhados residenciais [39]. Estes dispositivos são fixos, entretanto a rede não possui restrições quanto ao fornecimento de energia aos dispositivos. A rede possui capacidade de auto-organização, em que os equipamentos podem realizar roteamento, e sem restrição de escalabilidade. Este modelo de rede é similar às redes *peer-to-peer* de compartilhamento (*gnutella, napster, kazaa, freenet, metanet, waste e etc*).

Redes Ad Hoc em Malha possuem a propriedade de autoconfiguração, em que os dispositivos da rede automaticamente estabelecem a conexão na forma de redes ad hoc e mantêm a conectividade. Como em redes ad hoc não existe um elemento central e controlador, todos os dispositivos podem trocar informações diretamente entre si, ponto a ponto. A topologia é constituída por três níveis ou camadas: Internet ou rede externa; *backbone*; clientes [40]. A Figura 2.7 apresenta um modelo de arquitetura de redes Ad Hoc em Malha, contendo os principais componentes e os níveis de dispositivos. Na prática, uma rede Ad Hoc em Malha é híbrida, permitindo a utilização em conjunto de várias tecnologias de rede: cabeada ou sem fio.

Internet ou rede externa estabelece uma conexão com a rede, exterior ao ambiente dos usuários. Roteadores de borda, ou servidores que realizam esta tarefa podendo utilizar enlaces de rede sem fio. Poderá haver mais de um circuito para conexão externa [41].

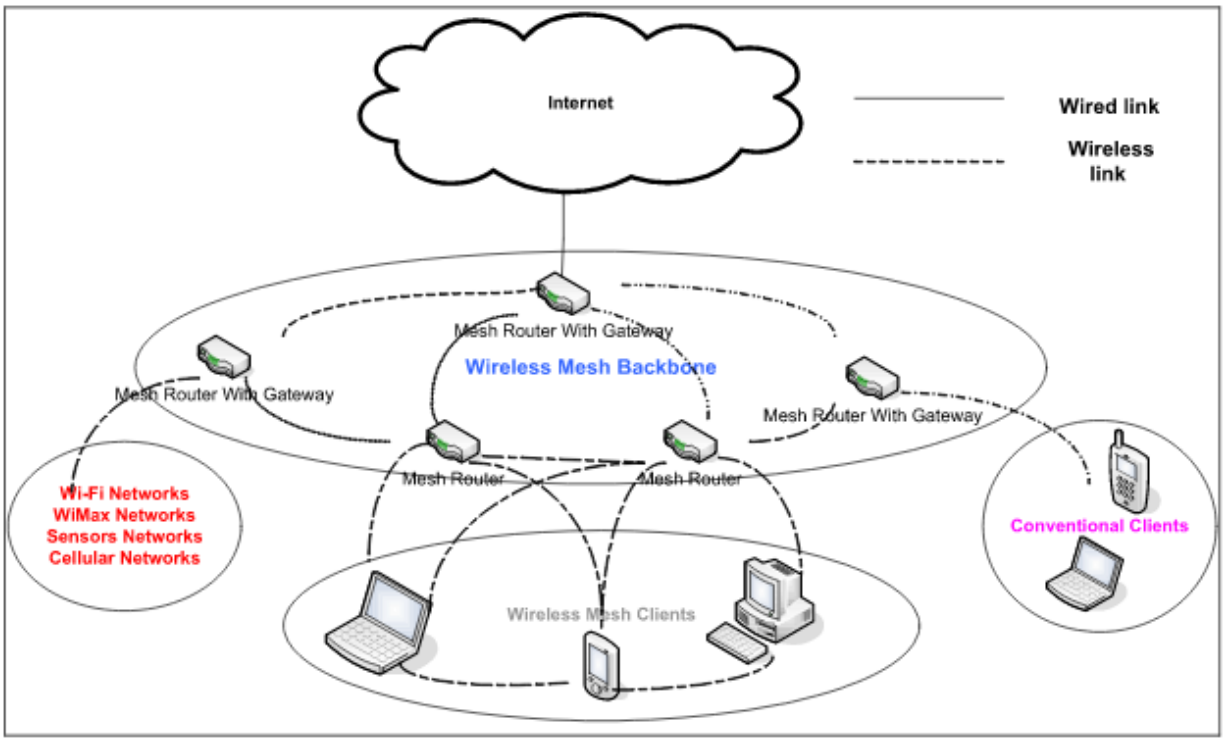


Figura 2.7: *A hybrid wireless mesh architecture* [4].

Backbone é um conjunto de dispositivos que possuem a funcionalidade de roteadores, que formam a infraestrutura principal da rede, ou seja, é a espinha dorsal da rede Ad Hoc em Malha. Esta camada comporta todos os dispositivos que realizam a função de *gateway* ou roteamento e constroem uma malha com topologia parcialmente ligada ou completa. Na maioria das vezes os dispositivos da rede possuem pouca ou nenhuma mobilidade, fixados em determinados locais [41].

Clientes são o conjunto de dispositivos utilizados pelo usuário final. Podem possuir mais de uma interface de rádio, e estas podem conectar-se a mais de um elemento concentrador. Não é necessária capacidade de roteamento. Permite uma variedade de equipamentos, desde sensores, estação móvel celular, *notebook*,ou seja, qualquer dispositivo de rede sem fio [41].

As redes WMN's utilizam enlaces de rádio, fazendo que seu desempenho seja baseado no desempenho das redes sem fio. Analogamente, as dificuldades da topologia de rede sem fio também são herdados pela rede Ad Hoc em Malha. Há diversos motivos para a utilização destas redes, sendo que os principais são [40]:

- Confiabilidade: cada dispositivo é um transmissor intermediário que encaminha os pacotes até seu destino final. Estes possuem capacidade de entrar e sair da rede. Cada dispositivo é capaz de, dinamicamente, mudar seu padrão de encaminhamento baseado em sua vizinhança. Assim, a topologia em malha aperfeiçoa a confiabilidade, pois resultará na transmissão dos pacotes via enlace alternativo até seu destino;

- Auto-organização: com a aprendizagem dinâmica de rotas e atribuição de endereços IP, os dispositivos possuem capacidade para conectar a seus vizinhos. Caso ocorra falha ou remoção de um nó, a topologia poderá ser restaurada de tal forma que os serviços não sejam interrompidos;

- Escalabilidade: os dispositivos podem entrar e sair da rede enquanto executam uma aplicação compatível com outros nós da rede. Este aspecto permite estender a área de cobertura de uma rede Ad Hoc em Malha, com a alocação de novos dispositivos em localizações apropriadas que possam se comunicar com os nós existentes da rede. Entretanto, a quantidade de dispositivos possíveis de serem inseridos é o limite máximo referente ao número total de nós que possa ter determinada rede sem degradar os serviços em função do número de saltos.

Há desvantagens deste tipo de rede, que são todas as dificuldades apresentadas pelas redes sem fio Ad Hoc, exceto o fornecimento de energia, visto que os dispositivos estão fixos.

2.3.3 Protocolos de Roteamento

O IETF (*Internet Engineering Task Force*), é um grupo informal de caráter internacional aberto, composto por técnicos, agências, fabricantes e pesquisadores, o qual se destina ao desenvolvimento de padrões para a Internet. Este construiu um grupo de trabalho que tem por finalidade discutir os problemas e padronizar um ou mais algoritmos de roteamento para uma rede sem fio Ad Hoc [42].

O algoritmo de roteamento é a parte do software da camada de rede responsável pela decisão sobre a linha de saída a ser usada na transmissão dos pacotes de entrada, ou seja, rotear pacotes de uma máquina de origem para uma ou mais máquinas de destino.

Um dos maiores desafios em redes sem fio Ad Hoc refere-se à possibilidade de perda de comunicação, seja por interferências ou por mobilidade dos seus dispositivos. Diante disto, a função de rotear mensagens é de fundamental importância e possui influência direta no desempenho de toda a rede. Os protocolos de roteamento utilizam informações imprecisas sobre o estado dos enlaces da rede, afetando o rendimento do protocolo e consequentemente o desempenho da rede.

Uma outra dificuldade surge no momento em que deseja enviar uma mensagem para outro dispositivo, pois é necessário encontrar o nó de destino para então definir uma rota entre os mesmos e mantê-la até o fim da comunicação tornando necessário um mecanismo de manutenção de rotas.

Essas dificuldades estão sendo focos de pesquisas. Diversos protocolos de roteamento e várias formas de manutenção de rotas são propostos. Estes protocolos no entanto devem lidar com limitações típicas desses tipos de rede como consumo de energia dos dispositivos móveis, banda passante limitada e altas taxas de erro. Em suma, os protocolos de roteamento ad-hoc dividem em dois grupos [43]: *Table-driven* e *On-demand*.

Os protocolos pertencentes ao grupo *table-driven* são aqueles que utilizam tabelas de roteamento para manter a consistência das informações de roteamento em todos os dispositivos. Desta forma incluem-se os seguintes protocolos:

- DSDV (*Destination-Sequenced Distance-Vector Routing*);
- WRP (*Wireless Routing Protocol*);
- CGSR (*Clusterhead Gateway Switch Routing*).

Os protocolos do grupo *on-demand* são aqueles com a característica de criar rotas somente quando desejado por um nó fonte. Assim, incluem-se os protocolos:

- AODV (*Ad Hoc On-Demand Distance Vector Routing*);
- DSR (*Dynamic Source Routing*);
- LMR (*Lightweight Mobile Routing*);
- TORA (*Temporally Ordered Routing Algorithm*);
- ABR (*Associativity-Based Routing*);
- SSR (*Signal Stability Routing*).

No decorrer desta revisão bibliográfica não é possível identificar um consenso a respeito do melhor protocolo de roteamento ad-hoc, pois cada protocolo possui vantagens e desvantagens de acordo com situações específicas. Segundo o grupo de trabalho do IETF[42], existe uma lista de qualidades favoráveis para os protocolos de roteamento em redes ad-hoc. Estas qualidades são descritas em [43].

Abaixo apresentam-se algumas métricas definidas para a avaliação de protocolos de roteamento:

- Vazão e atraso fim-a-fim;
- Tempo de aquisição de rota: fundamental para os algoritmos de roteamento que estabelecem rotas sob demanda;
- Porcentagem de pacotes entregues fora de ordem;
- Eficiência: algumas medidas podem ser obtidas para verificar a eficiência de um protocolo de roteamento. Como exemplo pode-se obter o número médio de bits de dados transmitidos por bits de dados entregues. O objetivo é verificar a eficiência na entrega de dados dentro da rede. Outro exemplo é o número médio de bits de controle transmitidos por bits de dados entregues. Neste caso, verifica-se qual o *overhead* causado pela parte de controle do algoritmo de roteamento.

As redes Ad Hoc em Malha possuem características que as diferenciam de outros modelos, mesmo sendo considerada uma particularidade de uma rede Ad Hoc [39]:

- Topologia: as redes em malha possuem *backbone* sem mobilidade, ou nós com pouca mobilidade;
- Tráfego: a predominância do tráfego é entre os nós móveis e o *gateway* da rede;
- Interferência entre caminhos: existe a possibilidade de interferência de sinais entre os nós que fazem parte da rede e outras redes em funcionamento;
- Diversidade de canais: a rede beneficia-se da possibilidade de utilização de diversidade de canais no processo de comunicação com outros nós.

A Tabela 2.3 apresenta uma comparação dos aspectos fundamentais entre os principais modelos de redes de computadores.

Tabela 2.3: Comparação entre Modelos de Redes de Computadores

Característica	**Cabeada**	**MANET**	**Rede de Sensores**	**Rede em Malha**
Topologia	Estática	Móvel	Estática	Estática
Tendência de Tráfego	Qualquer par de nós	Qualquer par de nós	Sensor ao sink	Nó móvel ao *gateway*
Interferências entre Caminhos	Não	Sim	Sim	Sim
Capacidade do Enlace	Fixa	Variável	Variável	Variável
Diversidade de Canais	Não Aplicável	Não	Não	Sim

Em [35] são descritos vários tipos de protocolos de roteamento ad-hoc, suas características, qualidades e problemas. A tabela 2.4 apresenta uma comparação em relação a alguns dos protocolos de roteamento para redes sem fio Ad Hoc.

Tabela 2.4: Comparação de Alguns Protocolos Para Redes Sem Fio Ad Hoc

Protocolo	**Tipo**	**Arquitetura**	**Métrica**	**Suporte a QoS**	**Número da RFC**
OLSR	Pró-Ativo	Plano	Saltos	Não	3626
TBRPF	Pró-Ativo	Plano	Saltos	Não	3684
AODV	Reativo	Plano	Saltos	Não	3561
DSR	Reativo	Plano	Saltos	Não	4728
CEDAR	Híbrido	Hierárquico	QoS	Sim	Não
ZRP	Híbrido	Hierárquico	Saltos	Não	Não

2.4 FRAGILIDADES DAS REDES AD HOC

Diversas vulnerabilidades encontradas nas arquiteturas das redes tradicionais, também são possíveis nas redes sem fio Ad Hoc. Isto se deve às características destas redes que enfatizam tais vulnerabilidades, possibilitando novas formas de explorá-las. Além do que, as redes Ad Hoc possuem fragilidades que são próprias e que não estão presentes em outras arquiteturas de redes [14, 44]. As propriedades especiais das redes Ad Hoc que dão maior ênfase nas vulnerabilidades já conhecidas por outras arquiteturas de rede ou que propiciam novas fragilidades específicas em redes Ad Hoc, distinguem-se [14]:

- A natureza sem fio do serviço de enlace - os dispositivos são capazes de monitorar a utilização da rede por dispositivos próximos que estejam dentro do alcance de seu receptor;
- O modelo de comunicação descentralizado ou ponto-a-ponto - os dispositivos são capazes de se comunicarem diretamente, uns com os outros;
- A mobilidade - a topologia de rede muda dinamicamente;
- O modelo colaborativo de comunicação - os dispositivos dependem uns dos outros para estabelecimento e manutenção da conectividade na rede;
- O uso frequente de fontes de energia que se extinguem com o uso - os dispositivos móveis usam fontes de energia portáteis.

Estas propriedades fazem com que as redes sem fio Ad Hoc se tornem mais vulneráveis que as redes estruturadas, possibilitando uma larga visão de ataques, por exemplo: Escuta passiva, personificação ou spoofing (uma entidade assume a identidade de outra) e negação de serviço. Os atacantes estão aptos a explorá-las com o objetivo de :

- Escutar promiscuamente transmissões orieundas de dispositivos próximos;
- Comunicar-se diretamente com qualquer dispositivo que esteja dentro de seu alcance de transmissão;
- Mover-se para coletar informações sobre a atividade de dispositivos que estão distantes ou para escapar da monitoração de dispositivos próximos;
- Praticar a não-colaboração, com o intuito de economizar sua própria bateria ou para provocar disfunções no encaminhamento de pacotes na rede;
- Provocar a realização de atividades desnecessárias com objetivo de acelerar a exaustão das fontes de alimentação de outros dispositivos.

Além disto, nas redes estruturadas, serviços de roteamento e autoconfiguração são realizados por dispositivos projetados para esta finalidade e também de segurança, denominados de roteadores e sevidores de autoconfiguração. Estes dispositivos, no entanto, executam um conjunto

planejado de funções, possuindo uma localização especial na topologia da rede, consequentemente com uma proteção cuidadosa, tanto física quanto lógica. Desta maneira, estes dispositivos possuem baixa vulnerabilidade, pois não existem funções genéricas, que possibilitam a desativação de funções desnecessárias. Também, para enfatizar esta baixa vulnerabilidade há a facilidade de ativar proteções nestes dispositivos, tanto lógicas ou físicas, em relação ao seu respectivo posicionamento dentro da rede. Este posicionamento dos dispositivos se dá normalmente em pontos de concentração dentro de partes controladas da rede [14, 44].

Entretanto, em redes Ad Hoc, os serviços básicos juntamente como os demais serviços de rede são dotados de maneira descentralizada e com participação total de todos os dispositivos da rede. Os dispositivos são implementados, constantemente, em equipamentos de computação móveis que podem possuir hardware e software genéricos, os quais estão sujeitos a diversas vulnerabilidades em relação ao sistema operacional, defeitos de softwares, porta de fundos(*backdoors*), vírus, dentre outras [14, 44].

Assim, de acordo com [14, 44], de uma forma mais constante haverá dispositivos em redes Ad Hoc com funcionamento ruins ou comprometido. Neste sentido, os dispositivos incorretos podem estar realizando algum ataque a rede Ad Hoc e ainda estar se movendo dentro da mesma, com a finalidade de realizar ataques a outros dispositivos ou escapando da monitoração de seus vizinhos. Este aspecto dificulta a detecção dos ataques e a identificação do(s) dispositivo(s) incorreto(s) por nós corretos na rede.

2.4.1 Principais Ataques a Redes Ad Hoc

As vulnerabilidades intrínsecas das redes sem fio são encontradas nas Redes Sem Fio Ad Hoc. Também, uma rede Ad Hoc possui vulnerabilidades específicas em função da tecnologia associada, principalmente a ausência de infraestrutura e ao encaminhamento colaborativo de mensagens.

O roteamento exige colaboração distribuída dos dispositivos destas redes. Estes dispositivos no entanto estão sob o controle de usuários da rede, e não de administradores. Esta particularidade facilita a criação de ataques, em que o objetivo é explorar vulnerabilidades dos algoritmos cooperativos. A ausência de mecanismos centralizados impossibilita o emprego de técnicas usuais de autenticação [45].

Os enlaces de rádio estão sujeitos a ataques passivos até interferências ativas. Diferentemente de redes com infraestrutura tradicionais, em que o atacante precisa ter acesso físico à rede e passar por vários mecanismos de defesas, como *firewall* e *gateways*, as redes Ad Hoc utilizam o ar como meio de transmissão, possibilitando que o atacante tenha acesso diretamente a rede [35].

Os principais ataques às redes Ad Hoc podem ser divididos em dois grupos básicos: passivos e ativos. Os ataques de característica passiva não interferem no funcionamento da rede, sendo caracterizados pela interceptação dos dados sem alteração dos mesmos. Estes ataques são difíceis de serem identificados. A espionagem de dados é feita através do meio inseguro com objetivo

de roubar as informações dos usuários ou descoberta de elementos da rede, como por exemplo, a topologia. Para minimizar este problema, pode-se utilizar recursos de outras camadas, protocolos que fazem uso de criptografia (HTTPS, SSH ou IPSec) [45, 2].

Os ataques ativos caracterizam-se pela criação, alteração ou descarte de dados em trânsito pelo atacante. Esta classe de ataques pode atuar em diferentes camadas do modelo RM-OSI. Os atacantes podem ainda ser internos ou externos. Ataques internos são aqueles que fazem parte da rede e se passam por membros da mesma, sendo que em alguns casos podem ser usuários autênticos da rede. Já ataques externos são formados por grupos que influenciam, mas não participam da rede.

O atacante, no entanto, faz uso de um ou vários dispositivos que são configurados para utilizar a mesma faixa de frequência das vítimas, com o intuito de provocar erros nas transmissões. O administrador do ambiente de rede deverá observar qual dispositivo possui o nível de sinal constante, e assim evitar rotas para o mesmo. Outra opção é fazer uso de analisador de espectro para encontrar o dispositivo malicioso. O espalhamento espectral aumenta a tolerância do sistema a interferências. Uma variação deste tipo de ataque é a interferência esporádica. O atacante utiliza o mesmo procedimento, mas em períodos aleatórios, em que as vítimas fazem maior uso de retransmissões, que também provoca o aumento do consumo da bateria. Assim, o usuário tem sensação de lentidão na rede. Por tratar-se de evento esporádico, a detecção é difícil.

Uma rede sem fio Ad Hoc possui dispositivos que na maioria das vezes possuem restrições de energia, como *notebooks*, celulares, *tablets*, sensores dentre outros. O ataque por exaustão de bateria tem como objetivo consumir os recursos de energia através de geração de retransmissão continuamente, com modificações maliciosas na camada de enlace. Este método pode ser aplicado diretamente à vítima, no momento de iniciar uma transmissão, sendo que o atacante gera pacotes com objetivo de causar colisão. Na colisão, a vítima será obrigada e retransmitir o quadro. O tratamento para este ataque é difícil, pois envolve modificações na camada de enlace [46].

Os dispositivos maliciosos podem gerar problemas nos protocolos de roteamento, seja com *loops*, rotas falsas, caminhos não ótimos ou até mesmo encaminhamento seletivo. Esse tipo de ataque é de difícil detecção pois, para dispositivos móveis, o funcionamento está correto, embora, de fato, esteja apresentando anomalias [45, 2].

Ataque bizantino recebe esta nomenclatura em função dos fatos históricos relatados sobre generais dos exércitos bizantinos. Os generais, distribuídos em campo de batalha, com suas tropas tinham a missão de organizar ataques aos inimigos. A comunicação entre os generais era realizada por mensagens. Mas existiam generais traidores e, com objetivos de confundir os demais, modificavam as mensagens com alteração de datas e horários dos ataques planejados. No ataque bizantino, um conjunto de dispositivos tem procedimentos semelhantes aos generais corruptos, realizando a alteração de alguns dados dos pacotes transmitidos na rede. A fim de garantir a veracidade das informações, faz-se uso de assinatura digital. Outras formas de soluções são baseadas em roteamento seguro, ou outros que tentam minimizar as retransmissões, como proposta apresentada por [47].

Os ataques de estouro da tabela de roteamento têm seu principio em protocolos ad hoc próativos, que armazenam as rotas anunciadas pelos nós vizinhos. Um nó malicioso publica rotas para nós não existentes. O atacante divulga um grande conjunto de rotas falsas com objetivo de aumentar a tabela de roteamento dos nós, até que seja estourada [48]. As redes Ad Hoc, com dispositivos com recursos reduzidos, como por exemplo, sensores, *notebooks*, *tablets*, celulares, devido ao número elevado de mensagens, podem ser bastante prejudicadas. Pode-se limitar o número máximo de valores na tabela de roteamento para solucionar este problema.

O ataque de replicação de pacotes tem a finalidade de ocupar o meio de transmissão. O atacante utiliza o encaminhamento de cópias de pacotes de roteamento antigo. Para contornar este problema, faz uso do número de sequência para identificar que sejam inseridos, na rede, pacotes que não são válidos. Entretanto, este tipo de ataque consome recursos do meio de transmissão e do processamento de dispositivos.

O ataque da pressa explora a maneira como os protocolos reativos constroem a tabela de roteamento. Quando um atacante recebe uma mensagem RREQ (*Route Request*), que é um *frame* de pedido de rota que é encaminhado a todos dispositivos que estão ao alcance de determinado dispositivo [49], o atacante responde antes dos demais dispositivos da rede, de forma que a rota para o destino passe pelo dispositivo malicioso. Os protocolos reativos armazenam, na tabela de roteamento, somente a primeira resposta recebida, neste caso, a do atacante, descartando as respostas dos outros dispositivos ficando o atacante com posição privilegiada na rede. A detecção deste ataque não é simples, pois o protocolo de roteamento considera este procedimento normal. Pode-se fazer uso de rotas múltiplas com objetivo de garantir que algumas rotas estariam funcionando. Uma proposta neste sentido é apresentada por [50] para uso em rede de sensores.

Também tem-se o ataque por direcionamento falso, que utiliza do mecanismo de funcionamento de pacotes *ECHO*. O atacante envia grande quantidade de mensagens, mas com pacotes modificados, objetivando redirecionar as respostas para um determinado dispositivo da rede. A vítima, identificada como o emissor das requisições *ECHO* receberá grande quantidade de respostas. A utilização de Sistema de Detecção de Intrusão poderá identificar esta anomalia, podendo no entanto fazer que o administrador seja capaz de bloquear o atacante.

Os protocolos que fazem uso de *HELLO* geralmente utilizam mensagens para se anunciarem aos vizinhos. No momento em que seus vizinhos recebem estes pacotes concluem que ambos estão dentro do alcance do enlace de rádio. No ataque denominado por inundação por *HELLO*, o atacante utiliza de amplificação de potência e encaminha as mensagens com informação sobre rotas boas para outras redes. As vítimas, por sua vez, atualizam as informações sobre as rotas, mas ao tentar utilizá-las não possuem êxito, pois o atacante está fora de alcance. A fim de evitar este ataque, os protocolos de roteamento verificam se os enlaces são bidirecionais.

O ataque por encaminhamento seletivo caracteriza-se pelo atacante não desejar prejudicar todos os dispositivos, mas alguns deles. É possível que o atacante realize a escolha de um determinado dispositivo e não faça o encaminhamento de seus pacotes. Este ataque, por sua vez, é difícil de ser detectado, devido a enlaces ruins ou com interferência. A utilização de rotas redundantes

minimizará o problema com a identificação do atacante. O caso extremo deste ataque é o descarte de todos os pacotes pelo atacante, denominado de buraco negro. Caso o atacante consiga atrair muito tráfego, a proporção deste ataque pode ser considerável, pois provocará consumo elevado de recursos dos dispositivos vizinhos. Diante disto um SDI deverá identificar este ataque na rede e então o administrador poderá bloquear o atacante.

O ataque tipo túnel de minhoca é caracterizado por dois dispositivos da rede combinarem e criarem um túnel com uso de enlace de baixa latência entre os mesmos. A principal finalidade é convencer os dispositivos da rede que estes podem se comunicar com determinados destinos, através de um único salto, em vez de utilizar múltiplos saltos da rede. Pode perceber que este ataque coloca os atacantes em posições privilegiadas e estes, por sua vez, poderão fazer uso deste privilégio quando assim o desejarem [51].

Os ataques por sequestro de sessão fazem uso de que muitas comunicações são protegidas somente no estabelecimento da sessão. No sequestro de sessões TCP, o atacante captura os dados transmitidos e recebidos pela vítima para determinar o número de sequencia utilizado. Realiza em seguida algum tipo de ataque de negação de serviço na vítima e continua a utilizar a sessão previamente estabelecida. A utilização de criptografia nas mensagens trocadas poderá inibir este ataque.

O ataque Sybil [52], é descrito em um primeiro momento para redes *peer to peer*, possuindo esta nomenclatura devido a um caso ocorrido nos Estados Unidos, em que uma mulher sofria de múltiplas personalidades. O ataque ocorre quando um único hardware assume várias identidades em uma rede. Em redes Ad Hoc, o atacante pode criar identidades falsas para rotas diferentes, objetivando centralizar várias rotas. O atacante, por sua vez, poderá fazer uso das várias identidades para ser privilegiado quando ocorrer votações na rede, pois alguns protocolos de roteamento utilizam de votação para escolha de dispositivos centrais. O atacante ainda poderá em redes que faz uso do princípio de confiabilidade, realizar ações de anomalias através de algumas de suas identidades. Para se defender deste ataque, realiza-se a validação da identidade dos dispositivos, presentes na rede, de acordo com seu endereço físico, através de dois métodos: validação direta e indireta. O método da validação direta direciona que cada dispositivo verifique se a identidade de outro é válida, enquanto que o método da validação indireta caracteriza-se por após o dispositivo ser validado, este poderá testemunhar ou refutar a validade de identidade de outro [52].

As conexões em portas TCP ou exploração de portas UDP podem ser utilizadas com objetivo de identificar quais serviços estão em execução. Podem ainda determinar o estado de escuta. Este ataque é chamado varredura de porta. Os principais tipos de varreduras são apresentados abaixo:

- Conexão TCP: o atacante conecta-se à porta do TCP da vítima e completa as três etapas da conexão: *flags SYN, SYN/ACK e ACK*. A identificação deste ataque pelo sistema alvo é relativamente fácil;
- TCP SYN: o atacante não realiza a conexão completa, em vez disso, envia apenas uma solicitação de conexão (*SYN*) para a vítima. Se a resposta for a confirmação do sistema alvo

(*SYN/ACK*), deduz-se que a porta está em estado de escuta. A conexão completa nunca é estabelecida, e os *buffers* alocados neste procedimento ficam ativos até o "*timeout*" definido pelo TCP for atingido. Este ataque, por sua vez, é de difícil detecção, comparando-se com a varredura completa do protocolo;

- TCP FIN: o atacante envia um segmento com o *flag FIN* ativado. A vítima, no entanto, deve responder com um segmento com o *flag RST* para cada porta fechada;
- Árvore de Natal (Xmas): o atacante envia um segmento com os *flags FIN, URG e PUSH* ativados. O sistema alvo deverá responder com um seguimento com o *flag RST* ativado para cada porta fechada;
- TCP NULL: Este ataque, é encaminhado um segmento com todos os *flags* desligados. O sistema alvo deverá responder com um segmento com o *flag RST* para cada porta fechada;
- TCP RPC: Neste ataque, são identificadas todas as portas que estão no estado aberto, e são inundadas com o comando *NULL* do sistema de Chamadas de Procedimentos Remoto com objetivo de identificar portas RPC. A partir disto pode-se obter a listagem dos programas e versões que estão em execução ligados a estas portas;
- Varredura UDP: devido à simplicidade do protocolo UDP, o atacante envia um pacote para cada porta que deseja verificar. Existem três resultados possíveis: Resposta de Porta UDP Inacessível, Resposta UDP, Sem Resposta. A Resposta de Porta UDP Inacessível indica que a porta solicitada está fechada, enquanto que a Resposta UDP indicará que a porta solicitada está aberta e pronta para uso, e por fim a Sem Resposta indicará que a porta solicitada pode estar aberta ou filtrada por *firewall*.

Esta tese com o objetivo de avaliar a proposta apresentada de um Sistema de Detecção e Classificação de anomalias em redes sem fio Ad Hoc, através de dados oriundos dos seguintes ataques de redes Ad Hoc:

- *EAPOLStart*: Uso do protocolo *Extensible Authentication Protocol (EAP)*, cujo objetivo é realizar um método de autenticação tanto na utilização do protocolo *Wired Equivalent Provacy (WEP)*, tanto para o protocolo *Wi-Fi Protected Access (WPA)*, em suas versões comerciais para acesso a redes Ad Hoc. Esta anomalia se caracteriza por uma carga excessiva de solicitação *EAPOL - Start*, que em um sobrecarregamento dos dispositivos da rede, responsável pela interconexão dos dispositivos da rede Ad Hoc;
- *BeaconFlood*: Solicitações do tipo gerenciamento, que têm a finalidade de transmitir milhões de *Beacons* não válidos, resultando na dificuldade que determinado disposito da rede Ad Hoc terá na identificação de um vizinho legítimo. Este quadro *Beacons* tem contribuição nos dispositivos na identificação da localização do IBSS(*Independent Basic Service Set*) de uma rede Ad Hoc [53];

- *Deauthentication*: Solicitações do tipo gerenciamento, que são injetadas na rede Ad Hoc. Os quadros pertencentes a esta anomalia são transmitidos como pedidos imaginários, os quais solicitam a desautenticação de um dispositivo que compõe a rede Ad Hoc [54];
- *RTSFlood*: Denominado *Request-to-Send Flood* é um quadro do tipo controle. Esta anomalia se baseia na transmissão em grande escala de pacotes ou frames RTS por um curto período de tempo. A inundação de frames RTS na Rede Wireless proporcionará o congestionamento na reserva do canal Wireless, resultando no processo de negação de serviço aos dispositivos da Rede Wireless [53].

Os ataques *Deauthentication e BeaconFlood* possuem como alvos os quadros de gerenciamento, enquanto que o *RTSFlood*, além dos quadros de gerenciamento também ataca quadros de dados, e o *EAPOLStart* visa os quadros de controle.

Os ataques de Desautenticação são utilizados empregando uma sequência de quadros falsos, em seguida, envia-se uma sequência ininterrupta de quadros de Desautenticação falsos.

Os *beacons* são quadros de sincronismo enviados periodicamente com informações sobre a rede. O seu emprego também possui a função de auxiliar sincronização na rede, geralmente são transmitidos a cada 100ms. Porém, no ataque de *BeaconFlood*, é produzido número demasiado de quadros com objetivo de impossibilitar clientes de se associarem ao ponto de acesso verdadeiro da rede.

Os quadros Requisição para Enviar – RTS são enviados pelos equipamentos de rede sem fio quando existem dados para serem transmitidos, e assim fazer a reserva do canal para comunicação. Porém, com objetivo de causar danos na rede, os quadros RTS utilizados no ataques possuem o campo duração alterado para um valor elevado, provocando a paralização da rede.

Os quadros EAPOL são empregados para transportar segmentos de rede oriundos do Protocolo Extensível de Autenticação – EAP sobre uma Rede Local – LAN, com objetivo de prover a comunicação entre um cliente (suplicante) e o ponto de acesso (autenticador). Ataques de EAPOL-Start geram excessivas requisições de inicializações de sessões EAPOL a um ponto de acesso, caracterizando-se como um Ataque de Negação de Serviço – DoS, com objetivo de paralisar o equipamento.

A Tabela 2.5 apresenta os principais ataques às redes Ad Hoc, sendo indicada também a camada do modelo RM-OSI em que podem ocorrer, e algumas mitigações.

A combinação destas anomalias caracteriza-se por ser um fator relevante que provavelmente indicará um possível ataque, sendo que a varredura de porta é uma ação precursora do ataque. A identificação do ataque pode ser realizada através de sistema de detecção de intrusos, e então o administrador poderá realizar ações de mitigação e prevenção.

Com a popularização da Internet, a família de protocolos TCP/IP tornou-se padrão em rede de computadores. Isso gerou problemas relacionados à segurança, pois em seu projeto original, essa não era a premissa principal destes protocolos. A preocupação do projeto era relacionada à

disponibilidade da rede, mesmo sob a falha de alguns dispositivos. Estes protocolos, aplicados diretamente em redes sem fio, herdam os problemas relativos à falta de segurança e também os problemas intrínsecos da tecnologia do IEEE 802.11, causados devido à utilização do meio comum de transmissão de radiofrequência.

Tabela 2.5: Principais Ataques às Redes Ad Hoc

Ataque	**Tipo**	**Camada**	**Mitigação**
Espionagem	P	Rede, Transporte ou Aplicação	Utilizar protocolos seguros como HTTPS, SSH ou IPSec
Interferência contínua ou esporádica	A	Física	Utilizar rotas alternativas ou espalhamento espectral
Exaustão de bateria	A	Enlace	Modificação da camada MAC
Ataque bizantino	A	Rede	Assinatura Digital
Estouro da tabela de roteamento	A	Rede	Limitar o tamanho da tabela de roteamento
Replicação de Pacotes	A	Rede	Fazer uso do número de sequencia
Ataque da pressa	A	Rede	Roteamento seguro
Direcionamento falso	A	Rede	Detecção com uso de SDI e bloqueio do nó pelo administrador
Inundação de HELLO	A	Rede	Utilizar roteamento com verificação de enlaces bidirecionais
Encaminhamento seletivo	A	Rede	Utilizar rotas redundantes
Buraco negro	A	Rede	Detecção com uso de SDI e bloqueio do nó pelo administrador
Túnel de minhoca	A	Rede	Roteamento seguro
Sincronização	A	Transporte	Limitar o número de solicitação de conexões
Sequestro de Sessão	A	Transporte	Utilização de criptografia
Sybil	A	Física, enlace e rede	Validar identidade através de endereço físico ou utilizar certificados digitais
Varredura de portas	A	Transporte	Detecção com uso de SDI e bloqueio do nó atacante pelo administrador
EAPOLStart	A	Método de autenticação.	Detecção com uso de SDI e bloqueio do nó atacante pelo administrador
BeaconFlood	A	Identificação da localização do BSS(*Basic Set Service*).	Detecção com uso de SDI e bloqueio do nó atacante pelo administrador
Deauthentication	A	Desautenticação de dispositivos da Rede.	Detecção com uso de SDI e bloqueio do nó atacante pelo administrador
RTSFlood	A	Negação de Serviço.	Detecção com uso de SDI e bloqueio do nó atacante pelo administrador

3 SISTEMAS DE DETECÇÃO DE INTRUSÃO

A popularização do acesso à Internet, fez com que muitas instituições utilizassem sua infraestrutura para realizar a comunicação entre seus diversos dispositivos. As demandas de negócios estão incentivando empresas e órgãos governamentais a desenvolver sofisticadas e complexas redes de comunicação e de informações. Essas redes possuem diversas tecnologias, incluindo armazenamento de dados, técnicas de criptografia e autenticação, voz e vídeo sobre IP, acesso remoto e sem fio, entre outros serviços. Entretanto, as redes corporativas estão se tornando mais acessíveis, sendo que organizações permitem que seus usuários utilizem os serviços da rede através de conexões oriundas da Internet [2, 55].

As motivações financeiras propiciam que os ataques que acontecem atualmente na web são bastante diferentes dos ataques tradicionais. Há uma constante evolução na inteligência destes ataques e com a dependência crescente da utilização da Internet, torna-se necessário a utilização de técnicas para garantir disponibilidade de serviços na rede. Também, são necessários recursos que garantam a integridade dos dados trafegados, bem como a identificação correta dos usuários remotos [2, 55].

Neste capítulo é apresentada a fundamentação teórica sobre Sistemas de Detecção de Intrusão, suas métricas de avaliação e sua aplicação em redes sem fio Ad Hoc.

3.1 SISTEMA DE DETECÇÃO DE INTRUSÃO

No Brasil, o CERT.br é o grupo de resposta à incidentes de segurança para a Internet, mantido pelo NIC.br, do Comitê Gestor da Internet no Brasil. O CERT.br é responsável por receber, analisar e responder a incidentes de segurança envolvendo redes conectadas à Internet no Brasil. A Figura 3.1 apresenta o número de incidentes reportados ao CERT.br nos últimos anos e percebe-se a ocorrência de aumentos significativos entre os anos, sendo o maior entre os anos de 2013 e 2014.

Estes incidentes podem ser mais numerosos, visto que parte dos incidentes não são reportados. De acordo com o CERT.br os incidentes de segurança mais frequentes no ano de 2017 foram a sondagem, responsável pela tentativa de identificar serviços ativos através de varreduras de portas, e ataques de *Denial os Service* - DoS, que têm o objetivo de tirar de operação um serviço, computador ou rede, através da utilização de um conjunto de computadores pelo atacante. Os principais incidentes recebidos pelo CERT.br no ano de 2017 são apresentados na Figura 3.2.

A segurança da informação é essencial para garantir o funcionamento correto das redes e sistemas computacionais. A segurança da informação está diretamente relacionada com as propriedades de integridade, confidencialidade e disponibilidade. A integridade fundamenta-se na

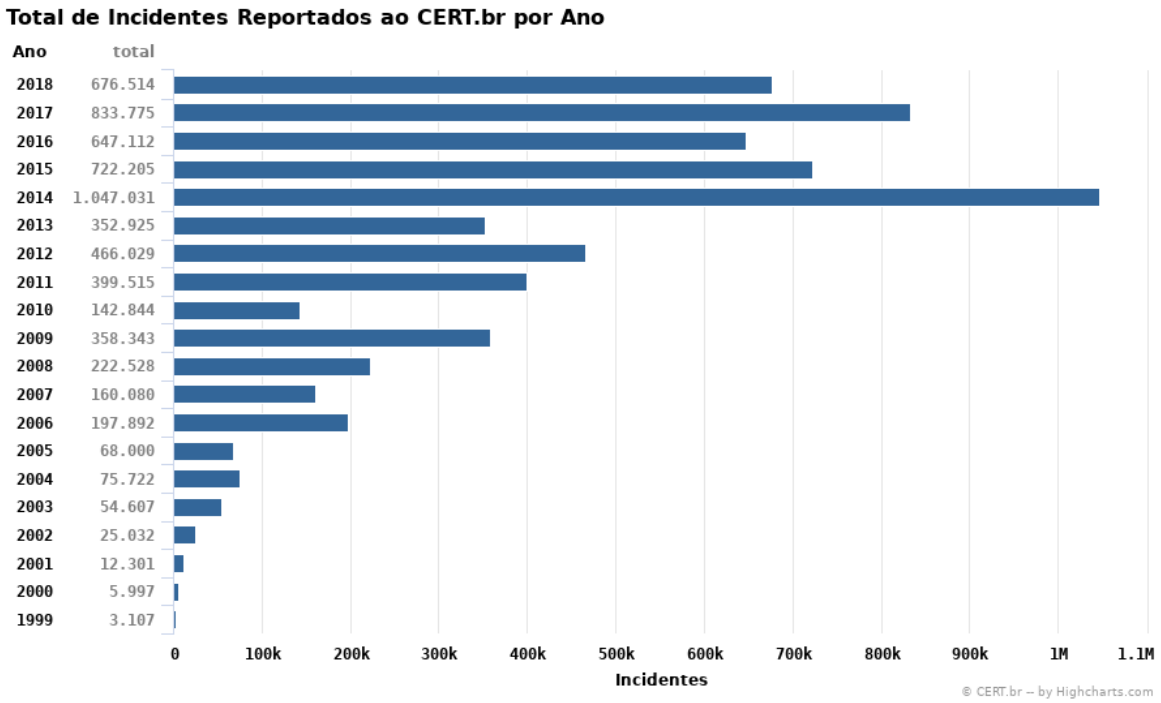


Figura 3.1: Incidentes Reportados ao CERT.br por Ano [1].

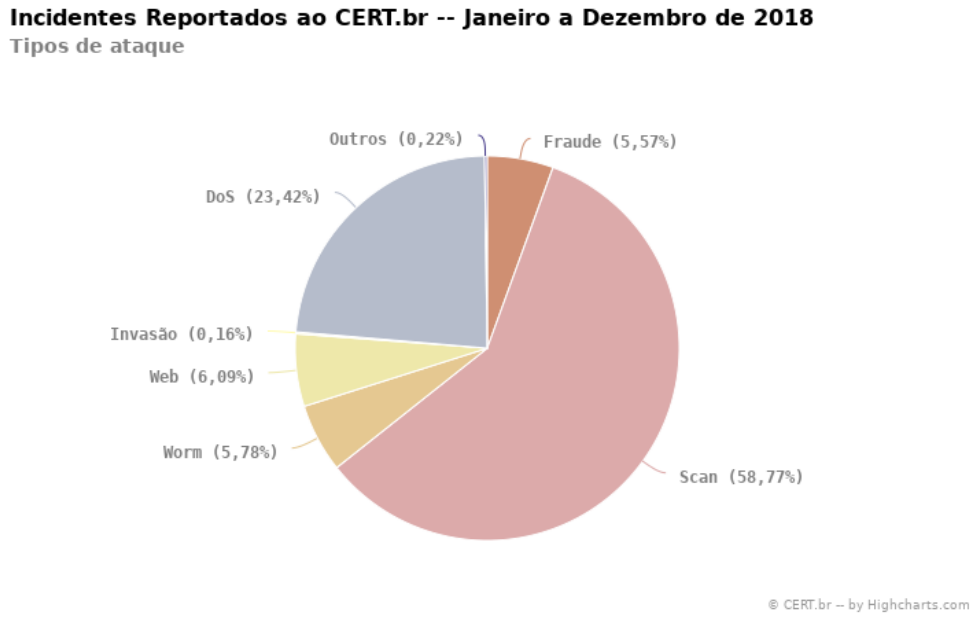


Figura 3.2: Incidentes Reportados ao CERT.br - Janeiro a Dezembro de 2017 [1].

exatidão da informação, a certeza de que a mesma não sofreu alterações. A confidencialidade é o procedimento que garante acesso apenas de usuários autorizados. A disponibilidade relaciona-se com a certeza de que os usuários permitidos terão acesso à informação quando necessitarem. As ações que tem o objetivo de comprometer estas propriedades de sistemas, ou redes, são classificadas como intrusão. Todos os procedimentos, técnicas capazes de identificar e isolar intrusos são denominados detecção de intrusão.

Sistema de Detecção de Intrusão - SDI, é um software utilizado para detectar acesso sem autorização a um computador ou rede. Este deve ser capaz de descobrir a existência de tráfego malicioso na rede de computadores. As anomalias detectadas pelo Sistema de Detecção de Intrusão podem ser: ataques à rede, exploração de vulnerabilidades de serviços, ataques à computadores, tentativa de aumento de privilégios, login sem autorização e tentativas de acesso a arquivos [2, 55].

De maneira geral a estrutura de determinado ataque é composta por três fases. A primeira obtêm informações sobre a vítima. A segunda fase é responsável pela busca de vulnerabilidades, sendo executada com o auxílio de softwares que identificam aplicações e suas fragilidades. A terceira fase destina-se a exploração das vulnerabilidades existentes, e a manutenção do acesso obtido na fase anterior [2, 56]. As modificações dos pacotes que trafegam na rede ocorrem a partir da segunda fase, podendo ser detectadas pelo Sistema de Detecção de Intrusão, pois já podem existir mudanças significativas na rede.

O Sistema de Detecção de Intrusão de forma geral possui um funcionamento dividido em três etapas:

1. Obtenção dos dados de auditoria: quando aplicados às redes existentes. Utilizam-se softwares como *tcpdump* para captura e leitura do tráfego na rede;
2. Seleção de características: os dados de auditoria formam um conjunto muito amplo e sua utilização direta acarreta custo elevado computacional. Estas informações são otimizadas afim de alcançar um conjunto de dados menor. Este novo subconjunto de dados possuem as características que melhor contribuem para o funcionamento do Sistema de Detecção de Intrusão, mantendo um nível aceitável de detecção. As redes com protocolos TCP/IP geralmente utilizam como dados os endereços de origem e destino, números de porta de destino e parte do conteúdo dos pacotes;
3. Análise: tarefa responsável pela avaliação que identificará que a rede está sob suspeita de ataque ou esta com funcionamento normal.

O SDI possui um conjunto de técnicas e métodos que são utilizados na varredura de atividades anômalas na rede, bem como no dispositivo local [2, 56]:

- Monitoramento e análise de atividades dos usuários e sistema;
- Gerenciamento de configuração de sistema e suas vulnerabilidades;
- Análise estatística de padrões baseados na comparação com ataques conhecidos;
- Análise de atividades anormais.

Há várias propostas para construção de Sistema de Detecção de Intrusão, como em aprendizagem de máquina [16, 17, 18, 19, 15, 20, 21] e mineração de dados [57]. O Sistema de Detecção de

Intrusão deve ser efetivo e eficiente. O Sistema de Detecção de Intrusão se torna efetivo quando consegue realizar a classificação correta de ações maliciosas ou normais. A propriedade de eficiência do Sistema de Detecção de Intrusão está na sua execução contínua, no entanto, com baixo consumo de memória e processamento dos dispositivos, não interferindo significativamente no desempenho da rede. Conforme Ferreira [2], a detecção de intrusão pressupõe que usuários e atividades de softwares são observáveis, fazendo com que quaisquer ações que o usuário ou aplicação inicia gerem atividades que podem ser gravadas em registros (*logs*), que são acessados pelo Sistema de Detecção de Intrusão. Estes registros ou *logs* são denominados "dados de auditoria". O SDI, então realiza análise nos dados de auditoria com objetivo de identificar comportamentos anormais de dispositivos da rede. Caso seja identificado um comportamento anômalo, o SDI poderá inferir que o sistema está sob ataque.

O procedimento de detecção de intrusão é classificado em duas categorias principais: detecção de assinatura e detecção de anomalia. A estratégia de detecção de assinatura fundamenta-se na identificação de padrões que correspondem ao tráfego de rede ou dados da aplicação e os compara com uma base de padrões (assinaturas) de anomalias conhecidas. Assim, anomalias conhecidas são detectadas com rapidez e baixa taxa de erro. Todavia, ataques desconhecidos não são detectados. A estratégia de detecção de anomalia esta fundada na construção de perfis de comportamento para o que será definido como atividade normal. Neste sentido, desvios da normalidade são reconhecidos como ameaças. Desta forma, os Sistemas de Detecção de Intrusão apoiados na detecção de anomalia têm a capacidade de adaptação a novas classes de anomalias, assim como, detectar ataques desconhecidos [2, 56]

Após a obtenção dos dados de auditoria, os Sistemas de Detecção de Intrusão podem realizar a análise dos dados de duas maneiras: *postmortem* e tempo real. A análise *postmortem* fundamenta-se na exploração de tráfego de dados como um único conjunto de dados, utilizando um procedimento mais rigoroso com um emprego maior de recursos computacionais. Este tipo de análise será útil para propósitos de engenharia de tráfego, análise de uso de recursos, criação de perfil de utilização, etc. Entretanto, outro tipo de análise de dados de auditoria é a análise em tempo real, que se destina a concentrar em analisar uma pequena janela de tráfego de dados, visando a fornecer alertas rápidos de anomalias do trafego. Esta técnica emprega procedimentos menos sofisticados devido à demanda de recursos para detecção de ataques [2].

Em relação a tecnologia usada para detecção e identificação da atividade suspeita, os Sistemas de Detecção de Intrusão são classificados em tipos baseados em assinatura e baseados em anomalia.

Os SDIs baseados em assinatura têm o princípio em um conjunto predefinido de padrões para detectar ataques. Os SDIs em assinaturas comparam os pacotes de dados com as assinaturas ou atributos de intrusões conhecidas para decidir se o tráfego observado é malicioso ou não. Essa abordagem é empregada apenas em ataques conhecidos. Este tipo de SDI utiliza um conjunto de regras para indicar intrusões observando eventos conhecidos e documentados. Este sistema está conectado a grandes bancos de dados, que armazenam ataques anteriores. Então, se o banco de

dados não estiver atualizado regularmente, haverá risco de não capturar o ataque. As definições de assinatura no banco de dados devem ser mais específicas, para que as variações de ataques conhecidos não sejam perdidas. Isso leva a um grande banco de dados, que pode armazenar muita memória no sistema. Os SDIs por assinatura são eficientes na detecção de intrusões conhecidas com assinaturas monomórficas. Entretanto, não são eficientes na detecção de intrusões ou intrusões desconhecidas com assinaturas polimórficas[19].

O SDI com base em anomalia opera no conceito de que o comportamento de ataque diverge do comportamento normal do perfil. Inicialmente identifica o perfil normal e, em seguida, o novo evento é comparado com o comportamento normal. Se a nova atividade se desvia do perfil normal, então é considerada anômala e gera um alarme. As variações entre o perfil normal e o recurso de monitoramento são analisadas usando várias técnicas, como análise estatística, aprendizado de máquina e técnicas de mineração de dados. Os sistemas de SDIs por anomalia sofrem com altas taxas de alarmes falso positivos e podem introduzir sobrecargas de processamento pesadas nos recursos de computação. Sua principal vantagem é a detecção de ataques desconhecidos. A detecção baseada em anomalia tem que ser adaptativa para poder enfrentar a mudança dinâmica da rede. Seu perfil normal deve representar a operação normal da rede. A mudança dinâmica deve ser incorporada imediatamente no perfil normal. A maioria dos SDIs atuais emprega as duas técnicas para obter melhor capacidade de detecção[19].

A análise dos dados de auditoria em Sistemas de Detecção de Intrusão pode ser realizada em dispositivos de forma individual ou por tráfego de rede. A construção dos Sistemas de Detecção de Intrusão consiste em agentes instalados nos dispositivos, que se comunicam com um sistema central, sendo então denominado *Host-based IDS*. A outra forma de análise de dados de auditoria por tráfego de rede ocorre quando os dados são obtidos através do tráfego da rede, sendo denominado *Network-based IDS*.

A Figura 3.3 apresenta uma arquitetura genérica de Sistema de Detecção de Intrusão, bem como, seus módulos que são descritos a seguir [2]:

- Captura de dados de auditoria: utilizada na fase de coleta. Os dados coletados nessa fase são analisados pelo algoritmo de detecção de intrusão para descobrir atividades suspeitas. Os dados coletados podem ser originados em *logs* de dispositivos ou de rede, registro de comandos ou *logs* de aplicações;

- Armazenamento: os dados de auditoria são armazenados, temporariamente ou definitivamente, para serem processados. Em alguns casos, o volume de armazenamento pode ser grande. Esta é uma característica importante no Sistema de Detecção de Intrusão, pois terá influência na eficiência do mesmo em emitir resposta caso encontre anomalias no arquivo *log*. Isto alavancou pesquisas da área com intuito de reduzir os dados de auditoria;

- Processamento ou detecção: o procedimento de processamento é o módulo mais importante do Sistema de Detecção de Intrusão. No processamento são executados algoritmos para encontrar provas (com certo grau de certeza) de comportamentos suspeitos nos dados de

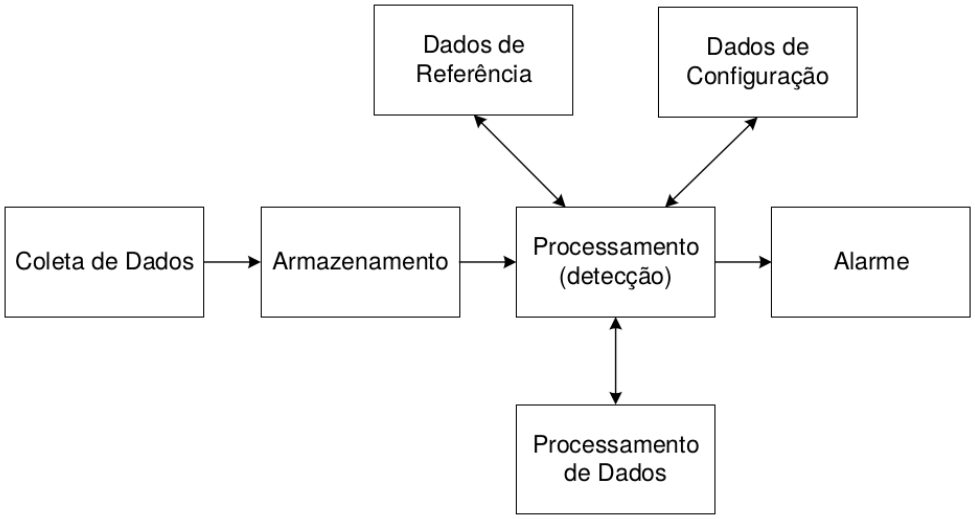


Figura 3.3: Arquitetura Geral de Sistema de Detecção de Intrusão [2].

auditoria;

- Dados de configuração: Este módulo afeta as operações do Sistema de Detecção de Intrusão. Possui parâmetros a respeito da localização dos dados de auditoria e de que forma serão realizadas repostas aos incidentes suspeitos. Por conseguinte, esta é a principal maneira de controle do responsável pela segurança sobre o SDI;
- Dados de referência: armazena informações sobre assinatura e/ou perfil normal de comportamentos conhecidos. No último caso, as atualizações do perfil considerado normal são permitidas e ocorrem em intervalos regulares;
- Processamento de dados: o processamento constantemente deve armazenar resultados intermediários, por exemplo, informações sobre assinaturas. Deve-se salientar que este epaço poderá aumentar, devendo ser gerenciado pelo administrador do Sistema de Detecção de Intrusão;
- Alarme: módulo do sistema responsável por informar a respeito da ocorrência de eventos suspeitos detectados pelo Sistema de Detecção de Intrusão.

A Tabela 3.1 apresenta a Matriz de Confusão para avaliação de detecção, que representam o grau de precisão de detecção [2]. Esta precisão de detecção são definidos a seguir:

- Falso negativo: incidentes intrusivos ou ataques, que são classificados pelo Sistema de Detecção de Intrusão como atividades normais;
- Falso positivo: ocorrência de eventos normais que são classificados pelo SDI como atividades intrusivas;
- Verdadeiro negativo: atividade normal classificada corretamente pelo SDI;

- Verdadeiro positivo: atividade intrusiva classificada corretamente pelo SDI.

Tabela 3.1: Matriz de Confusão de Avaliação de Sistema de Detecção de Intrusão

Tipo de Evento	**Classificação pelo SDI - Normal**	**Classificação pelo SDI - Ataque**
Normal	Verdadeiro Negativo	Falso Positivo
Ataque	Falso Negativo	Verdadeiro Positivo

A partir da Matriz de Confusão, calculam-se certas métricas de desempenho. A acurácia define o grau de exatidão ou precisão demonstrado pela Matriz de Confusão. A acurácia global é a precisão das detecções corretas obtidas pelo Sistema de Detecção de Intrusão em relação ao total de classificações. Esta acurácia global é calculada conforme mostra a Equação 3.1

$$PrecisoGlobal(OA) = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Em que:

TP é a taxa de verdadeiro positivo;

TN é a taxa de verdadeiro negativo;

FP é a taxa de falso positivo;

FN é a taxa de falso negativo.

São apresentados até o momento conceitos a respeito de segurança da informação e sua importância na atualidade. Com o crescimento das redes *wireless* e uso da Internet, houve um crescimento de usuários dos serviços providos pelas redes. A segurança da informação tornou-se imprescindível para disponibilização de serviços aos usuários.

Os Sistemas de Detecção de Intrusão tornaram-se ferramentas importantes para melhorar o nível de segurança das redes e das aplicações. Diversas técnicas são empregadas na tentativa de melhorar a taxa de detecção global, a exemplo de inteligência artificial, processamento digital de sinais, métodos estatísticos, séries temporais, entre outros.

O número de propostas para implementação de SDI são elevados, esse é um indício que se trata de uma área que ainda necessita de muitos estudos, principalmente para acompanhar o alto desenvolvimento das tentativas de ataques sem comprometer o funcionamento da rede ou dos sistemas.

3.2 SISTEMA DE DETECÇÃO DE INTRUSÃO EM REDES SEM FIO AD HOC

Ainda que as redes cabeadas e ad hoc sejam submetidas às mesmas vulnerabilidades, torna-se praticamente impossível a transposição direta dos modelos de Sistema de Detecção de Intrusão

utilizados nas redes cabeadas. Esta dificuldade decorre de diferenças nas características funcionais entre as duas redes.

Em redes cabeadas ou com infraestrutura é viável estabelecer claramente várias fronteiras de tráfego, permitindo a segmentação dos seus fluxos e desta maneira localizar detectores de intrusão pelo núcleo central de comunicação da infra-estrutura. Estes detectores de intrusão, no entanto, podem funcionar de forma localizada ou distribuída. Na estratégia distribuída detectores comunicam-se por meio de protocolos seguros, permitindo ao agregador central a interpretação mais ampla das diversas conexões entre os diferentes eventos de segurança.

As análises, bem como, as conclusões realizadas por Sistemas de Detecção de Intrusão acontecem de maneira rápida e estão fortemente integradas aos sistemas de gerenciamento, que possuem a funcionalidade de ativar alarmes aos administradores do SDI. Estes alarmes se originam por possíveis eventos considerados ameaçadores, sendo que a reação poderá ser de forma manual ou automática, sendo esta essencialmente composta por regras fixas de ações pré-determinadas.

Nas redes sem fio Ad Hoc, devido a sua natureza dinâmica e cooperativa torna-se difícil estabelecer de forma clara e duradoura limites físicos que permitam a utilização de tecnologias empregadas como firewalls, ou qualquer outro mecanismos centralizador de controle de tráfego.

Juntamente com estas características das redes Ad Hoc, os dispositivos das redes Ad Hoc podem ser submetidos a outras limitações como: limitações de energia, links de comunicação, largura de banda, capacidades reduzidas de armazenamento e processamento, possibilidade de operação descontínua de um nodo e comportamentos bizantinos no interor da rede.

Diante disto, é necessária a combinação destas características na construção de Sistemas de Detecção de Intrusão para redes Ad Hoc, tornando-os diferentes das abordagens utilizadas em redes com infraestrutura. Assim, espera-se que soluções de SDI para redes Ad Hoc fundamentam-se nas propriedades cooperantes e dinâmicas destas redes [58].

O Sistema de Detecção de Intrusão, conforme já mencionado é um componente de defesa fundamental em ambientes de redes de comunicação de Dados. Isto se deve, pois os mecanismos tradicionais de prevenção não são viáveis para a proteção de redes sem fio Ad Hoc. O SDI voltado para redes Ad Hoc possuem três componentes principais, que são: coleta, detecção e resposta de dados. A coleta de dados tem a funcionalidade de coleta e pré-processamento de dados, que consiste na transferência de dados para um formato comum, juntamente com o armazenamento de dados e encaminhamento para o módulo de detecção [59].

3.2.1 Arquitetura de Sistema de Detecção de Intrusão

As arquiteturas de Sistemas de Detecção de Intrusão para redes sem fio Ad Hoc são classificadas em três categorias, que são: *stand-alone* ou autônoma, cooperativa e hierárquica.

Na arquitetura *stand-alone* cada dispositivo da rede sem fio Ad Hoc é responsável por executar de maneira local o Sistema de Detecção de Intrusão, sendo que não há colaboração com os demais

dispositivos, mas há resposta local. Este tipo de arquitetura de SDI pode ter dificuldades em detectar ataques de rede [60].

Cooperativa é uma arquitetura na qual todos os dispositivos da rede sem fio Ad Hoc possuem seu próprio Sistema de Detecção de Intrusão local. Neste tipo de arquitetura os dispositivos da rede Ad Hoc decidem de maneira distribuída cooperativamente a respeito de um evento malicioso. Logo após a determinação de uma intrusão, os dispositivos compartilham essas informações, o grau de risco de ataque de ativos e tomam as ações necessárias para eliminar a intrusão usando precauções ativas ou passivas [60]. Conjuntamente, todos os dispositivos participam de uma tomada de decisão global de detecção, porém deve-se levar em consideração o tempo de atividade dos dispositivos.

Já a arquitetura hierárquica é uma abordagem multicamada, dividindo a rede sem fio Ad Hoc em *clusters*. Dispositivos específicos são selecionados, utilizando critérios específicos para atuarem como líderes de *cluster* e assumindo responsabilidades e funcionalidades no processo de detecção de intrusão, que são diferentes das dos demais dispositivos do *cluster*. Este tipo de arquitetura possui como vantagem principal o uso efetivo de recursos de restrição, mas possuem uma desvantagem para redes Ad Hoc, que é a propriedade de alta mobilidade desta rede, dificultando o estabelecimento de zonas e a escolha de dispositivos líderes em *clusters* [60].

O Sistema de Detecção de Intrusão possui um módulo denominado de mecanismo, que é responsável por detectar anomalias de maneira local utilizando os dados de auditoria coletados nas rede sem fio Ad Hoc. Este procedimento de detecção de intrusão local é realizado por algoritmos de classificação. Os algoritmos de classificação primeiramente realizam o pré-processamento nos dados de auditorias já rotulados, a fim de atender pré-requisitos dos algoritmos de classificação, bem como para garantir melhor eficiência no processo de classificação. Posteriormente, os algoritmos de classificação utilizando dados de treinamento realizam o processamento do classificador e, utilizam este classificador para testar os dados de auditoria local, com o intuito de classificar os dados de auditoria em "normal"ou "anormal"[59].

Em redes sem fio Ad Hoc há a presença de um método denominado *Watermarking* ou Marca d'água, que fundamenta-se na proteção dos dados relacionados que serão trocados entre os dispositivos da rede, imperceptíveis ao sinal de cobertura para transmitir os dados ocultos. Marcas d'água são aplicadas a fim de evitar a possível modificação dos mapas produzidos [59].

3.2.2 Detecção de Intrusão em Redes Sem Fio Ad Hoc Utilizando Algoritmos de Classificação

Sistemas de Detecção de Intrusão em redes sem fio Ad Hoc utilizam como norteador algoritmos de classificação supervisionada, ou seja, são algoritmos que fazem uso de uma base de treinamento rotulada para a classificação de dados de auditoria. Estes modelos de SDI utilizam uma arquitetura composta por agentes SDI locais, que têm a finalidade de detectar possíveis intrusões de maneira local.

Os algoritmos de classificação comumente utilizados nestes Sistemas de Detecção de Intrusão são: *MultiPayer Perceptron - MLP* [61], o modelo *Naive Bayes* [62], o modelo *Support Vector Machine - SVM* [63], o modelo Redes Bayesianas [59], o modelo J48 referente a Árvore de Decisão [64]. Todos esses modelos exigem dados de treinamento rotulados para sua criação. A coleta dos agentes SDI independentes forma o sistema SDI para o redes sem fio Ad Hoc. Cada agente SDI local é composto pelos seguintes componentes:

- Coletor de dados: é responsável por selecionar dados de auditoria locais e registros de atividades;
- Mecanismo de detecção de invasão: é responsável por detectar intrusões locais usando dados de auditoria local. A detecção de intrusão local é realizada usando um algoritmo de classificação.
- Mecanismo de Resposta: Se uma invasão for detectada pelo mecanismo de detecção, o mecanismo de resposta será ativado. O mecanismo de resposta é responsável por enviar um alarme local e global para notificar os dispositivos da rede sem fio Ad Hoc a respeito de um incidente de invasão.

Atualmente pode-se encontrar pesquisas que tenham o propósito de construir Sistemas de Detecção de Intrusão em redes sem fio, através da utilização de uma abordagem em camadas, mas ainda com a aplicação local em dispositivos de redes sem fio Ad Hoc [16, 17, 19, 15, 20, 21, 65, 66]. Esta abordagem se torna viável, pois tem como preocupação fundamental o processamento dos recursos, bem como sua duração de energia, sendo esta última ainda um problema para aplicação em redes Ad Hoc de forma geral. Diante disto observa-se a apresentação de propostas de Sistemas de Detecção de Intrusão utilizando algoritmos de classificação supervisionado e não-supervisionado, processamento digital de sinais, métodos estatísticos, séries temporais, entre outros para posterior classificação.

3.3 COMENTÁRIOS FINAIS

São apresentados nesse capítulo conceitos relacionados à segurança da informação e sua importância na atualidade. Com o crescimento das interligações das redes e uso da Internet, ocorre um aumento considerável de usuários dos serviços providos pelas redes. A segurança da informação tornou-se imprescindível para disponibilização de serviços aos usuários.

Os Sistemas de Detecção de Intrusão são ferramentas importantes para melhorar o nível de segurança das redes e das aplicações. Diversas técnicas são empregadas na tentativa de melhorar a taxa de detecção global, a exemplo de inteligência computacional, processamento digital de sinais, métodos estatísticos, séries temporais, entre outros.

O número de propostas para implementação de Sistemas de Detecção de Intrusão são elevados, esse é um indicativo que se trata de uma área que ainda necessita de muitos estudos, princi-

palmente para acompanhar o alto desenvolvimento das tentativas de ataques sem comprometer o funcionamento da rede ou dos sistemas.

4 INTELIGÊNCIA COMPUTACIONAL

Inteligência Computacional é uma área da computação e engenharias que investiga os princípios que tornam o comportamento inteligente possível. Estes comportamentos definidos e estudados são também chamados de *técnicas*, dentre as quais encontram-se: Redes Neurais Artificiais, Redes Bayesianas, Lógica Fuzzy, Algoritmos Evolucionários, Teoria de Jogos, Árvores de Decisão.

4.1 REDES NEURAIS ARTIFICIAIS

Os trabalhos relacionados às Redes Neurais Artificiais, também denominadas de Redes Neurais, têm se motivado pelo reconhecimento de que o cérebro humano consegue processar informações de maneira diferente em relação a um sistema computacional. Desta forma de acordo com [6] o cérebro é um sistema de processamento de informação altamente complexo, não-linear e paralelo.

Uma Rede Neural Artificial é um processador paralelamente distribuído, sendo composto por unidades de processamento simples. Esta Rede Neural possui a capacidade de armazenar conhecimentos adquiridos através de experiências e torná-los disponíveis para uso. As Redes Neurais Artificiais se assemelham ao cérebro humano pela intensidade das conexões entre os neurônios, que na Rede Neural Artificial é definida pelos pesos sinápticos, e são adaptados com a experiência adquirida [6]:

As Redes Neurais são compostas por neurônios, as quais estão conectadas de forma orientada. Outra propriedade importante das Redes Neurais é a presença dos pesos sinápticos para cada conexão presente em determinado neurônio. A Figura 4.1 apresenta o modelo de neurônio presente nas Redes Neurais, sendo composto por: sinais de entrada, pesos, função Soma, função de transferência e saída.

Estes componentes são apresentados a seguir:

- Entradas: é um conjunto de entradas ou elos de conexões orientadas. O sinal x_j presente na entrada da sinapse *j* conectada ao neurônio *k* é multiplicado pelo peso sináptico w_{kj}. O peso sináptico do primeiro índice *k* refere-se ao neurônio em questão e o índice *j* relaciona-se ao terminal de entrada da sinapse a qual o peso esta se referindo;
- Função Soma: Efetua a soma ponderada das entradas. A função soma pode ser representado pela Equação 4.1:

$$u_k = \sum w_{kj} x_j \tag{4.1}$$

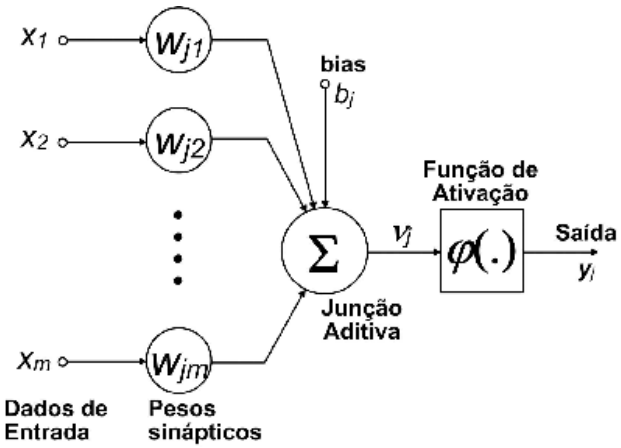


Figura 4.1: Modelo de Neurônio [5].

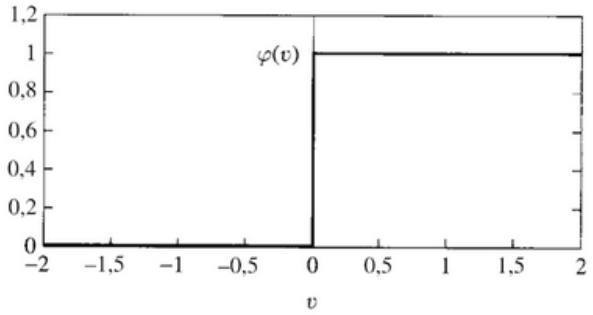


Figura 4.2: Função de Ativação de Limiar [6].

- Função de Ativação: Também denominada de Função de Transferência que mapeia a soma a uma saída final.

Funções de ativação comuns incluem a limiar e a sigmoide. A seguir será detalhado cada uma delas.

A função de ativação de limiar é definida na Equação 4.2. Assim, a saída de um respectivo neurônio será 1, caso a entrada for positiva e de forma análoga a saída será 0, caso a entrada seja negativa. A entrada da função de ativação é a função soma presente no modelo do neurônio. A Figura 4.2 mostra o gráfico da função de ativação de limiar.

$$\varphi(v) = \{1\, se\, v \geq 0;\, 0\, se\, v < 0\} \tag{4.2}$$

A função de ativação logística é a função mais utilizada em aplicações de Redes Neurais Artificiais, sendo composta por um gráfico na forma de *S*. Esta função é estritamente crescente. A função sigmóide é definida pela Equação 4.3, que esta representada graficamente na Figura 4.3.

$$\varphi(v) = \frac{1}{1 + exp(-av)} \tag{4.3}$$

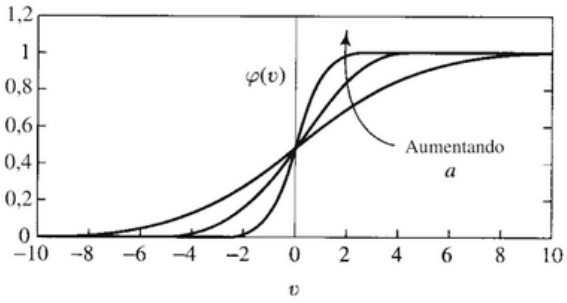


Figura 4.3: Função de Ativação Sigmóide [6].

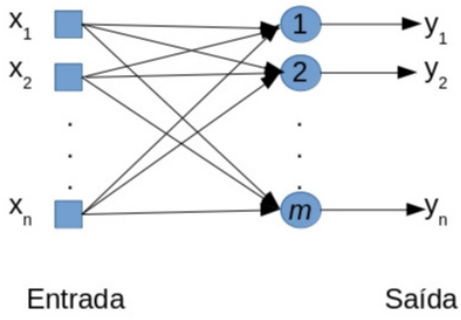


Figura 4.4: Redes Neurais de Camada Única.

Também há nas Redes Neurais a capacidade de adaptar os valores de seus pesos sinápticos, para otimizar performance. Esta característica de adaptabilidade dos pesos sinápticos tem a finalidade de fazer que as Redes Neurais Artificiais consigam atingir o melhor resultado possível na resolução do problema designado.

A implantação física das Redes Neurais possui tolerância a falhas. Desta forma, as Redes Neurais Artificiais implantadas fisicamente têm a capacidade de realizar processamento robusto, fazendo com que seu desempenho degrade, mediante condições de operações adversas [6].

Em se tratando de sua arquitetura, as Redes Neurais Artificiais podem ser classificadas em dois grupos distintos: as que não têm ligações recorrentes, chamadas aciclicas, e as cíclicas, que as possuem [67].

A Rede Neural de alimentação direta é organizada em camadas, sendo que determinada camada poderá receber apenas entradas de neurônios situados na camada imediatamente anterior. De acordo com [6], nas Redes Neurais Acíclicas poderão existir camadas especiais, denominadas de camadas escondidas ou ocultas, que não se conectam ao mundo exterior. As Figuras 4.4 e 4.5 apresentam os dois tipos de Redes Neurias Acíclicas, que são respectivamente denominadas de Redes Neurais de Camada Única e Redes Neurais de Várias Camadas [68].

As Redes Neurais Acíclicas de camada única, como a Rede *Perceptron* [67], são Redes Neurais que fundamenta-se na alimentação e possuem todas as suas entradas conectadas diretamente a uma determinada saída, conforme apresentado na Figura 4.4, sem a presença de camadas ocultas.

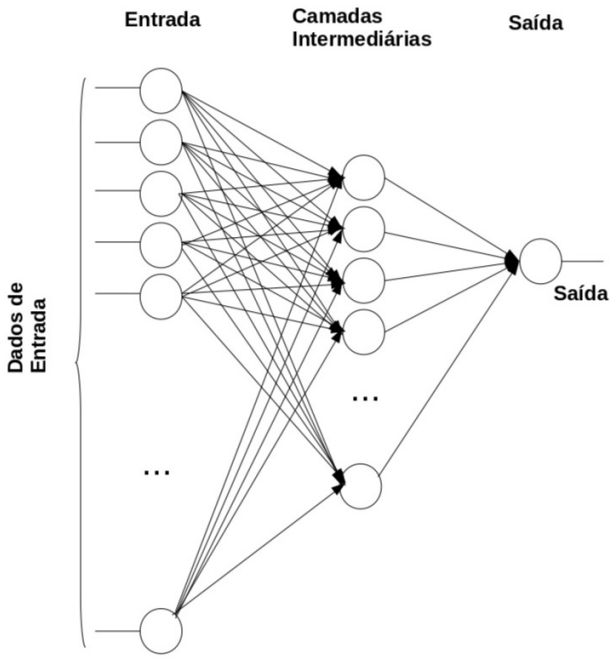


Figura 4.5: Redes Neurais de Várias Camadas.

Já as Redes Neurais Acíclicas de várias camadas possuem como principal aspecto a alimentação direta, mas com a presença de uma ou várias camadas ocultas entre a entrada e a saída, conforme apresentado na Figura 4.5. Redes Neurais com camadas ocultas têm maior poder computacional.

Redes Neurais Recorrentes, ou cíclicas, têm a característica de formar um sistema dinâmico, em que as saídas de determinada camada poderão realimentar entradas de outras camadas ou da própria camada em questão. Esta propriedade fará com que a Rede Neural possa representar dinâmicas complexas, inclusive caóticas, e exibir propriedade de memória.

Em Redes Neurais Recorrentes uma camada de saída pode realimentar todas as entradas dos demais neurônios de determinada Rede Neural, como também realimentar a sua própria camada de entrada [6]. Para as situações em que a camada de saída realimenta a sua própria entrada, ocorre o fenômeno chamado de auto-realimentação, que possui influência direta no processo de aprendizagem da Rede Neural, bem como em seu desempenho.

As redes recorrentes têm aplicação em processamento de dados sequenciais, como som, dados de séries temporais ou linguagem natural. As redes recorrentes possuem um *loop de feedback*, em que a saída do passo *n-1* é alimentada de volta à rede para afetar o resultado do passo *n*, e assim por diante para cada etapa subsequente. Por exemplo, se uma rede é exposta a uma palavra letra por letra, e é solicitado a adivinhar cada letra a seguir, a primeira letra de uma palavra ajudará a determinar o que uma rede recorrente pensa que a segunda letra pode ser. A Figura 4.6 mostra a estrutura de um Rede Neural Recorrente [7].

Uma característica das Redes Neurais Recorrentes é produzir modelos dinâmicos, que alteram ao longo do tempo, a fim de produzir classificações precisas dependentes do contexto dos

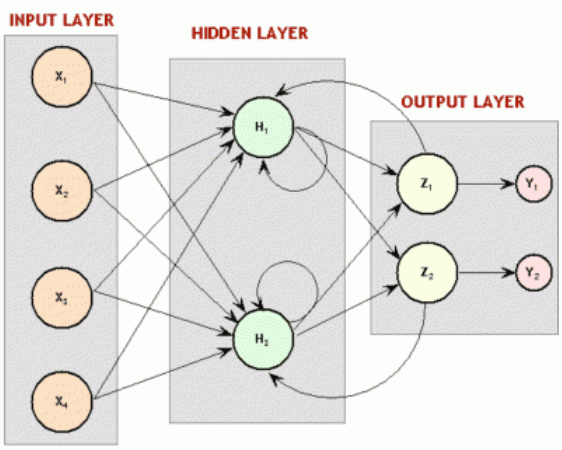


Figura 4.6: Rede Neural Recorrente [7].

exemplos que estão expostos. O modelo recorrente é eficiente pois inclui o estado oculto que determina a classificação anterior em uma série. Em cada etapa subsequente, esse estado oculto é combinado com os dados de entrada do novo passo para produzir: um novo estado oculto e, em seguida, uma nova classificação. Cada estado oculto é reciclado para produzir seu sucessor modificado. A Rede Recorrente que associa memórias e entrada remota no tempo é chamada de *Long Short-Term Memory*(LSTM) [7].

A utilização de Redes Recorrentes para ambientes de redes *wireless* e/ou redes sem fio Ad Hoc se torna desvantajoso, pois o seu processamento não é síncrono em relação as propriedades destas redes, principalmente em relação a mobilidade de seus dispositivos e, da manutenção da energia dos mesmos.

4.1.1 Aprendizagem em Redes Neurais

O processo de aprendizagem é um aspecto de Redes Neurais Artificiais que possibilita sua semelhança com o cérebro humano, sendo que para as Redes Neurais o nível da aprendizagem é garantido pela manipulação dos parâmetros, que são apresentados na Figura 4.1. De acordo com [6] o processo de aprendizagem das Redes Neurais Artificiais é definido como "o processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede esta inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre."

Para o aprendizado em Redes Neurais existem diversos algoritmos capazes de realizar a adaptação dos parâmetros, para que após uma quantidade finita de iterações possa convergir para uma solução viável. O critério de convergência para determinada solução esta subsidiado no processo de minimização de uma função objetivo ou na variação do erro de saída, sendo também possível a utilização de outros métodos de convergência [69]. O algoritmo de aprendizado visa à redução de uma função de custo, geralmente associada ao erro na saída do sistema.

O principal objetivo do processo de aprendizagem é encontrar os ajustes dos pesos sinápticos

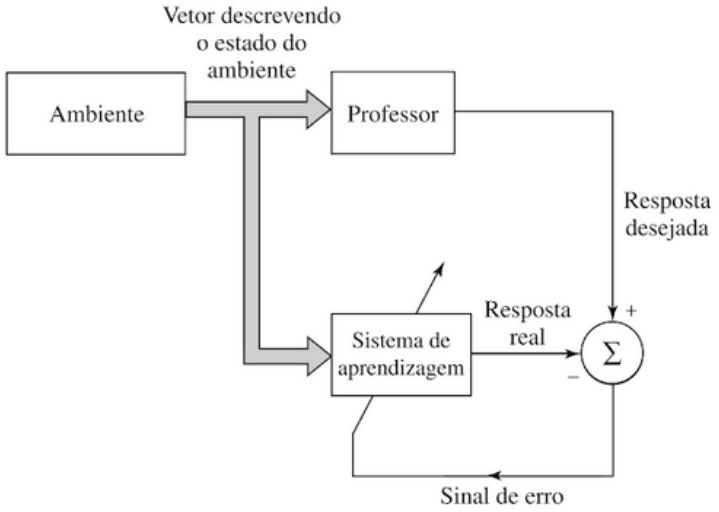


Figura 4.7: Diagrama em Blocos da Aprendizagem Supervisionada [6].

pertencentes a cada componente de entrada dos neurônios, possibilitando a convergência para o resultado esperado. Segundo [69] os algoritmos ou técnicas de aprendizagem aplicados as Redes Neurais Artificias podem ser classificados em três classes distintas:

- Aprendizado supervisionado;
- Aprendizado não-supervisionado;
- Aprendizado por reforço.

A aprendizagem supervisionada fundamenta-se em um conjunto de exemplos de entrada e saída. Conhecida a saída desejada para cada exemplo, pode-se calcular o erro e realizar o ajuste dos pesos, com o objetivo de aproximar a resposta encontrada da resposta desejada. A Figura 4.7 mostra um diagrama em blocos que representa o processo de aprendizagem supervisionada.

Este tipo de aprendizagem pressupõe o conhecimento do ambiente por um especialista, através de um conjuntos de entrada-saída. O ambiente é desconhecido para a Rede Neural Artificial. O especialista possui um conhecimento prévio do ambiente e fornece à Rede Neural a resposta desejada para o vetor de treinamento. Esta resposta desejada é a atividade ótima a ser realizada pela Rede Neural. Assim, as propriedades da Rede Neural são ajustadas através da combinação do vetor de treinamento com o sinal de erro gerado. O sinal de erro é definido como a diferença entre a resposta desejada e a resposta real da Rede Neural, sendo transferido à mesma através de treinamento [6]. Após o aprendizado, a Rede Neural estará apta a gerar a melhor solução para um determinado conjunto de informações, sem a presença do especialista.

A aprendizagem não-supervisionada consiste em efetuar o aprendizado de Rede Neural Artificial através do processamento de um conjunto de informações, sendo que não leva-se em consideração a presença de um especialista ou professor, ou seja, para este tipo de aprendizado não se tem o conhecimento das saídas desejadas para as entradas participantes durante o treinamento. Desta

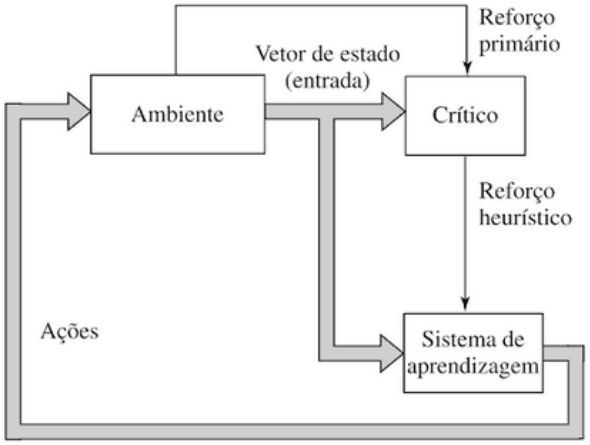


Figura 4.8: Diagrama em Blocos da Aprendizagem Por Reforço [6].

maneira os ajustes dos pesos sinápticos pertencentes em cada entrada são realizados tomando como base os valores de entrada [69].

O aprendizado não-supervisionado atende tipicamente às Redes Neurais Artificiais que estão sendo aplicadas em problemas em que há a necessidade de se categorizar determinados dados, tendo como fonte as informações de entrada. De acordo com [6], após as Redes Neurais se adaptarem às regularidades estatísticas dos dados de entrada, estas por sua vez têm a capacidade de codificar as características destes dados identificando novas classes, como por exemplo na Aprendizagem Competitiva. No caso da aprendizagem competitiva os neurônios da Rede Neural competem entre si para responder a características dos dados de entrada para depois classificar os mesmos.

A aprendizagem por reforço consiste no aprendizado de um mapeamento de entrada-saída, através da iteração contínua com o ambiente, objetivando maximizar uma recompensa. A Figura 4.8 mostra o diagrama em blocos da aprendizagem por reforço, tendo como base a ação de um agente crítico que tem a funcionalidade de converter o sinal de reforço primário, oriundo do ambiente, em um sinal de melhor qualidade, denominado sinal de reforço heurístico. Este tipo de aprendizagem fundamenta-se na aprendizagem por reforço atrasado, sendo que o sistema observa uma sequência temporal de estímulos, também oriundos do ambiente, os quais resultarão em sinais de reforço heurísticos. O objetivo principal desta aprendizagem é minimizar a função de custo para avançar, que é definida como a expectativa do custo cumulativo de ações tomadas no decorrer de uma sequências de passos. A máquina de aprendizagem tem a responsabilidade de identificar ações tomadas ao longo deste tempo, que sejam determinantes para o comportamento global do sistema, for fim realimentá-las no ambiente [6].

A Figura 4.9 demonstra o comportamento de um neurônio no processo de aprendizado supervisionado, tendo como elementos fundamentais o vetor de entrada, camada de neurônios ocultos, neurônio de saída e função somatória.

O neurônio tem uma saída $Y_k(n)$, que é o resultado da função de ativação do neurônio. Na

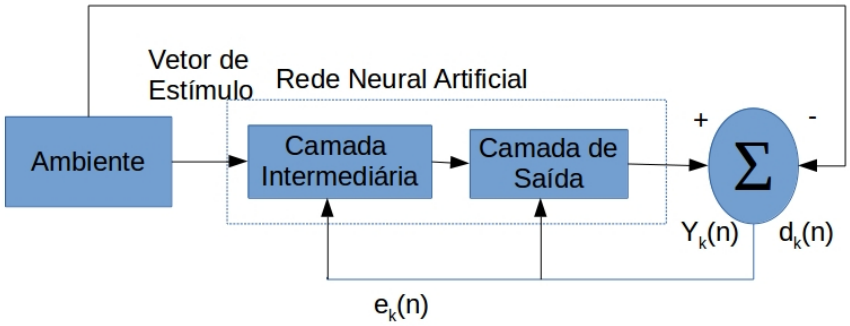


Figura 4.9: Aprendizagem Por Correção de Erro.

Figura 4.9 há a presença de uma saída desejada $d_k(n)$, que possivelmente será diferente da saída $Y_k(n)$ do neurônio. Desta forma, observa-se que a subtração entre a saída gerada pelo neurônio $Y_k(n)$ e a saída desejada $d_k(n)$ tenha como resultado um sinal de erro $e_k(n)$, que será retornado ao neurônio, possibilitando o ajuste dos seus respectivos pesos da camada de entrada, sendo geradas novas saídas, conforme é definido nas equações abaixo:

- Um padrão é apresentado e o sinal propagado através da rede. A saída linear de cada neurônio é dada pela equação 4.4 [70] :

$$v = \sum_{i=1}^{m} w_i x_i + b \tag{4.4}$$

onde:

w_i : i-ésimo peso sináptico;

x_i : i-ésima entrada;

b : peso correspondente ao "bias";

m : número total de entradas.

A saída não linear de cada neurônio é representada pela equação 4.5 [70]:

$$y_j(n) = \acute{\varphi}_j(v_j(n)) \tag{4.5}$$

onde:

$y_j(n)$: saída do *j*-ésimo neurônio na iteração *n*;

$\acute{\varphi}_j$: função não-linear;

- O erro da Rede Neural na camada de saída da *n*-ésima iteração é dado pela equação 4.6 [70].

$$e_j(n) = d_j(n) - y_j(n) \tag{4.6}$$

onde:

d_j : *j*-ésima saída desejada;

y_j : *j*-ésima saída calculada pela Rede Neural.

- É calculado o gradiente local da última camada através do erro gerado na camada de saída e da derivada do erro utilizando a equação 4.7 [70].

$$\delta_j(n) = e_j(n)\acute{\varphi}_j(v_j(n)) \tag{4.7}$$

- Propagação do Erro: O gradiente local de cada neurônio das camadas anteriores é calculado através da equação 4.8 [70].

$$\delta_j(n) = \acute{\varphi}_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \tag{4.8}$$

onde:

j : índice do neurônio da camada atual;

k : índice do neurônio da camada imediatamente posterior.

- Ajuste dos pesos sinápticos: Após calculado cada gradiente local o ajuste dos pesos sinápticos é dado pela equação 4.9 [70]:

$$\Delta w_{ji}(n) = \eta \varphi_j(n) y_j(n) \tag{4.9}$$

onde:

η : taxa de aprendizagem da rede.

- A adaptação dos pesos é realizada pela equação 4.10 [70].

$$w_{ji}^{k+1}(n) = w_{ji}^{k}(n) + \Delta w_{ji}(n) \tag{4.10}$$

- Para cada padrão apresentado à Rede Neural é medido o erro instantâneo na equação 4.11 [70].

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in C} ej^2(n) \tag{4.11}$$

onde:

C : Conjunto de todos os neurônios da camada de saída.

O processo de Aprendizagem por Correção de Erro é implementado por Redes Neurais Artificias através da utilização de algoritmos capazes de realizar diversos ciclos de aprendizagens

em tempos consideráveis úteis, denominados de épocas. Neste sentido um dos principais algoritmos utilizados em Redes Neurais Supervisionadas é o *Multilayer Perceptron - MLP*, também conhecido como *Perceptron* de Múltiplas Camadas, sendo também um dos mais utilizados em problemas com a necessidade de classificação [6].

O algoritmo *Multilayer Perceptron* é composto de sua entrada, saída e de pelo menos uma camada oculta entre a entrada e a saída. As camadas ocultas por sua vez, são as camadas que não possuem entradas e nem saídas, conforme visto antes e apresentado na Figura 4.5. Este tipo de Rede Neural esta sendo utilizada em larga escala para a resolução de problemas complexos, pois possui como propriedade fundamental o treinamento supervisionado, como por exemplo o algoritmo de retropropagação do erro ou *backpropagation* [6].

A Rede Neural Artificial do tipo *Multilayer Perceptron* possuem três aspectos básicos para o seu funcionamento:

- Os seus neurônios possuem uma função de ativação não-linear;
- A Rede Neural possui pelo menos uma camada oculta;
- A Rede Neural tem o aspecto de ser fortemente conectada, ou seja, cada um dos neurônios possuem ligação com todos os neurônios da camada anterior.

A unidade básica para Redes Neurais *Multilayer Perceptron* é formada pelo neurônio, sendo composto por um combinador linear e uma função de ativação. Todo neurônio pertencente a Rede Neural *Multilayer Perceptron* tem como propriedade receber contribuições dos respectivos neurônios da camada anterior, através do seu combinador linear de saída, que corresponde a soma dos pesos de entrada de determinado neurônio. Estas Redes Neurais, por sua vez, não possuem a técnica de aprendizado supervisionado, sendo necessário acoplar um algoritmo de aprendizado, como por exemplo o algoritmo de retropropagação do erro(*backpropagation*).

As Redes Neurais Artificiais em sua maioria não possuem o conhecimento prévio dos valores corretos de seus pesos sinápticos para que possa atingir a solução correta do problema em questão, levando em consideração um conjunto de entrada. A partir disso, percebe-se a necessidade da aplicação de um aprendizado da Rede Neural, o qual permitirá que esta obtenha uma solução a mais próxima possível da desejada, fazendo uso da manipulação de seus pesos sinápticos.

O algoritmo de retropropagação do erro, é utilizado em Redes Neurais Artificiais com arquitetura supervisionada com a finalidade de ajustar os seus pesos sinápticos. Outro aspecto do algoritmo *backpropagation* é além de ajustar os pesos sinápticos da Rede Neural, também a cada ciclo de aprendizagem ter a possibilidade de realizar a minimização de erros entre a saída gerada pelos neurônios e a saída desejada, ou seja, diminuir a taxa de erro a cada ciclo de aprendizagem [71].

O processo de propagação do retorno do erro, ou seja, a retropropagação do erro da saída para as camadas ocultas das Redes Neurais é realizado através da obtenção da derivação do erro

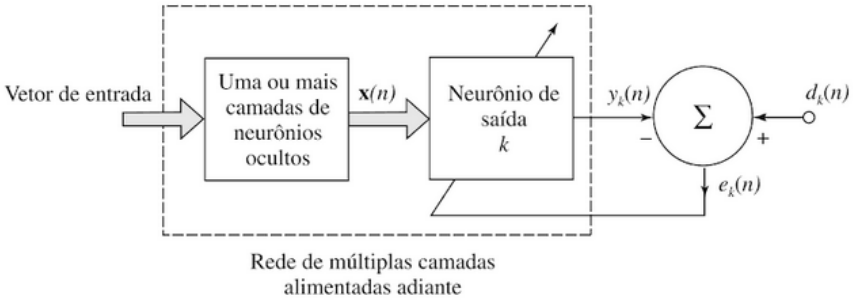


Figura 4.10: Diagrama em Blocos - *Backpropagation* [6].

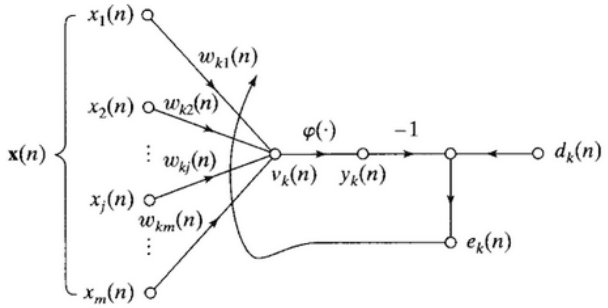


Figura 4.11: Fluxo do Sinal do Neurônio de Saída - *Backpropagation* [6].

global da Rede Neural e sua aplicação nos ajustes dos pesos. Com isto, cada neurônio pertencente a camada oculta receberá apenas uma porção do sinal de erro total, sinal este obtido pela diferenciação da saída desejada pela saída obtida, conforme é apresentado na Figura 4.9.

Este processo de retornar para cada neurônio apenas uma porção do sinal de erro total, que é proporcional à contribuição relativa de cada neurônio na formação da saída original, será repetido de camada por camada, até que cada neurônio da Rede Neural *Perceptron* receba o seu sinal de erro, o qual representa sua contribuição relativa para o erro total. Assim, cada neurônio com participação na saída original terá seus pesos sinápticos atualizados, permitindo a convergência da Rede Neural para atingir uma apróximação da solução desejada, possibilitando a identificação de padrões existentes nas informações de entrada. A Figura 4.10 apresenta o diagrama em blocos da aprendizagem por correção de erro, enquanto que a Figura 4.11 apresenta o fluxo do sinal do neurônio de saída para correção de erro.

Para os algoritmos de Aprendizagem por Correção de Erro percebe-se a existência de ciclos de aprendizados, sendo atualizados os respectivos pesos sinápticos de cada neurônio pertencente a Rede Neural *Perceptron*. A fim de fazer que estes algoritmos retornem uma solução e não fiquem executando continuamente, se tornando úteis para solucionar diversos problemas em várias áreas, há a necessidade de se estabelecer critérios de paradas. Em relação ao algoritmo de aprendizado por correção de erro, *backpropagation*, os critérios de parada são:

- Finalizar o treinamento após n ciclos;

- Finalizar o treinamento após o erro quadrático médio;
- Ficar abaixo de uma constante α;
- Finalizar o treinamento quando a porcentagem de classificações corretas estiver acima de uma constante α;
- Combinação dos métodos acima.

A Rede Neural do tipo *Multilayer Perceptron* é utilizada na proposta desta tese para a classificação dos dados oriundos de redes sem fio Ad Hoc, sendo que esta Rede Neural encontra-se na segunda etapa do sistema proposto. A definição de sua utilização fundamenta-se em alguns fatores, dentre os quais esta a presença de sua utilização em trabalhos relacionados à classificação de dados em redes *wireless* e redes Ad Hoc [16, 19, 20, 21, 72, 73], sobretudo com taxa de classificação viável.

Outro aspecto que contribuiu para a utilização de Rede Neural *Multilayer Perceptron* é a boa resposta do algoritmo em relação a aplicação às bases de dados utilizadas nesta tese para classificação dos dados obtidos. As bases utilizadas referem-se a uma base de redes sem fio Ad Hoc coletadas em ambiente acadêmico [8] e outra é um conjunto de dados disponibilizados publicamente pelo laboratório Lincoln do MIT denominada KDD 99 [74], sendo comumente utilizada para validação de Sistemas de Detecção de Intrusão. Estes dados mostram que tráfego de redes sem fio Ad Hoc são adaptáveis ao algoritmo, possibilitando atingir um melhor resultado na classificação, contribuindo com os administradores dessas redes.

4.1.2 Mapas Auto-Organizáveis

As Redes Neurais Artificiais denominadas de Mapas Auto-Organizáveis (MAO), também denominadas de Redes de *Kohonen*, possuem como objetivo principal a elaboração de sistemas que se organizem internamente a partir da distribuição dos dados de entrada, entretanto sem a presença de um especialista do problema a ser resolvido. As Redes de *Kohonen* têm como fonte inspiradora o fato de que as informações no cérebro humano são organizadas de forma espacial. As áreas do córtex formam "mapas" de espaços sensoriais referentes a neurônios responsáveis por respostas específicas a determinados estímulos de certas regiões, como por exemplo respostas específicas para frequências nas áreas do cérebro destinadas a audição e visão. Neste sentido as Redes de *Kohonen* geram aglomerados com alta atividade de resposta a um determinado estímulo.

Kohonen [75] descreve a Rede Neural Mapa Auto-Organizável (MAO), uma técnica de aprendizagem não-supervisionada bastante utilizada na visualização de dados. Esta rede é muito difundida por conseguir representar um espaço multidimensional em um espaço de baixa dimensão, algumas vezes apenas duas dimensões, denominado de mapa. Um mapa consiste de neurônios que têm associados a eles uma posição no mapa e vetores peso de mesma dimensão dos vetores de entrada. Para formar o mapa, os vetores peso competem entre si para saber qual será ativado

para cada vetor de entrada, a fim de serem ajustados. Abaixo é apresentado o algoritmo que descreve o funcionamento da Rede Neural Mapa Auto-Organizável, desde o pré-processamento até a classificação final dos dados de entrada [76].

```
 1 --- Fase 1 : pré-processamento
    2 Lê bases de dados;
    3 Filtra valores inválidos;
4 --- Fase 2 : Treinamento
    5 Parametriza a rede
    6 Dimensiona a Camada de Saída
    7 Enquanto existirem entradas e não tiver sido alcançado qualquer critério de parada
        8 Recebe entrada para treinamento
        9 Normaliza a entrada
        10 Calcula distâncias euclidianas
        11 Identifica o neurônio vencedor
        12 Ajusta peso do neurônio vencedor
        13 Ajusta peso dos vizinhos do neurônio vencedor
    14 Fim do Enquanto
15 --- Fase 3 : Classificação
    16 Enquanto existirem entradas normalizadas
        17 Recebe entrada normalizada para classificação
        18 Calcula distâncias euclidianas
        19 Classifica a entrada, identificando o neurônio vencedor
    20 Fim do Enquanto
```

Abaixo é apresentada a estruturação das Redes de *Kohonen* ou Mapas Auto-Organizáveis:

- MAO é composto por uma camada única, onde encontra-se a entrada I e a saída U;
- A entrada da Rede de *Kohonen* corresponde a um vetor no espaço d-dimensional em $\mathbb{R}$, o qual será representado por $x_k = [\xi_1, \ldots, \xi_d]^T$, *K* = 1, . . . , *n*;
- Cada neurônio *j* de saída possui um vetor denominado *w*, também representado no espaço $\mathbb{R}$ associado ao vetor de entrada $x_k, w_j = [w_{j1}, \ldots, w_{jd}]^T$;
- Os neurônios estão interconectados através da relação de vizinhança apresentada na estrutura do mapa.

Os neurônios de entrada recebem o mesmo valor no momento de inicialização das Redes *Kohonen*, tendo como função de ativação destes neurônios a função identidade. A função identidade é calculada para cada neurônio, através da soma ponderada das entradas, a qual para o MAO é a saída. Os neurônios que melhor se alinham ao vetor de entrada obterão melhor resposta e vencerão a disputa. Para o neurônio vencedor será atribuída uma determinada vizinhança, que será permitido a aprendizagem, através da adaptação de seus pesos seguindo a Equação 4.12. Em seguida será realizada um processo de renormalização com o objetivo de que todos os vetores de entradas dos neurônios escolhidos tenham o mesmo módulo.

$$w\left[t+1\right] = w\left[t\right] + \alpha\left[t\right] h\left[t\right] \left(x\left[t\right] - w\left[t\right]\right) \tag{4.12}$$

onde:

$\alpha\left[t\right]$: taxa de aprendizado;

$h\left[t\right]$: é a função que determina a vizinhança entre o neurônio vencedor e seus vizinhos, segundo a distância Euclideana.

A Equação 4.12 indica uma aproximação entre o vetor de entrada x_k e o vetor de saída w_j . Em uma próxima apresentação para este mesmo neurônio de entrada a probabilidade de se ter a melhor resposta de alinhamento e ser o neurônio escolhido é alta. Portanto neurônios se tornam especialistas em determinadas regiões do espaço de entrada das Redes de *Kohonen*.

Este processo descrito anteriormente é repetido para todos os vetores de entrada, sendo repetido diversas vezes. Já o tamanho da vizinhança à qual será permitido o processo de aprendizado se inicia com um valor elevado a fim de trabalhar com toda a Rede de *Kohonen* e diminuindo com o tempo, bem como a taxa de aprendizagem α presente na Equação 4.12 .

Segundo [6] os neurônios de saída das Redes de *Kohonen* representam as possíveis saídas do mapa, independente dos neurônios presentes na entrada, apesar de cada saída estar conectada a todos neurônios de entrada. Assim, o neurônio com maior estímulo será ativado, tornando os demais inibidos. Entretanto a posição do neurônio na saída será ativada através da menor distância Euclidiana, comparando-se com as demais posições.

Levando em consideração o estado de ativação da posição *i* da matriz do Mapa de *Kohonen* em relação ao estímulo do vetor de entrada x_k afirma-se que:

1. $y(x_k)$ = 1, se $i(x_k) = arg_j min \left\|x_k - w_j\right\|$;
2. $y(x_k)$ = 0, caso contrário.

Em que:

- $i(x_k)$: É a posição i da matriz do Mapa de *Kohonen*;
- $\|.\|$: A medida da distância, através da norma Euclidiana, $1 \geq j \geq N$. Sendo N a quantidade de neurônios existentes na saída;
- $y(x_k)$: Informa o estado de ativação da posição i da matriz do Mapa de *Kohonen* em relação ao estímulo do vetor de entrada x_k.

Em Redes de *Kohonen* um processo importante após a definição do neurônio vencedor em determinado instante, é a identificação de sua vizinhança. Esta vizinhança será determinada pelos neurônios mais próximos do neurônio vencedor, pois voltam-se a possuir uma excitação maior pelo fato de conter uma distância lateral menor. Em relação à determinação desta vizinhança

[6] afirma “O neurônio vencedor localiza o centro de uma vizinhança topológica de neurônios cooperativos”.

Outro aspecto no processo de determinação da vizinhança é garantir uma similaridade entre os neurônios integrandes de uma determinada região, sendo esta região composta pelo neurônio vencedor e seus vizinhos. De acordo com [75] para garantir a similaridade deve-se aplicar os ajustes dos pesos, conforme apresentado na Equação 4.12, tanto no neurônio vencedor, tanto em sua vizinhança. Assim, permite-se que o mapa organize-se geograficamente, pois os neurônios que não pertencem a vizinhança não terão seus pesos ajustados.

Supondo que h_{ij} refere-se à vizinhança tendo como centro o neurônio vencedor i, e que cada neurônio pertencente a vizinhança seja representado por j, e que d_{ij} defina a distância lateral entre o neurônio vencedor i e determinado neurônio pertencente a vizinhança j, será necessário atingir as seguintes condições:

1. A vizinhança topológica h_{ij} é simétrica em relação ao ponto máximo definido por d_{ij} = 0, ou seja, para o neurônio vencedor i a distância d_{ij} é zero;
2. A amplitude da vizinhança topológica h_{ij} decresce monotonamente com o aumento da distância lateral d_{ij}, decaindo a zero para $d_{ij} \rightarrow \infty$. Esta é uma condição necessária para a convergência, significando afirmar que quão mais longe estiver o neurônio vizinho do vencedor, menor será a sua ativação.

A Rede Neural Mapa Auto-Organizável - MAO também é avaliada para contribuir com a proposta apresentada nesta tese, a fim de agrupar os dados oriundos de redes sem fio Ad Hoc para posterior classificação. O processo de agrupamento dos dados da rede Ad Hoc encontra-se na primeira etapa do sistema proposto, logo após a obtenção dos dados da rede. A escolha de avaliar a Rede Neural Mapa Auto-Organizável - MAO consiste no aspecto da presença de sua utilização em trabalhos relacionados a redes *wireless* e/ou redes Ad Hoc [77, 78, 79, 80].

Outro aspecto analisado para a utilização da técnica de inteligência computacional para o agrupamento dos dados de redes Ad Hoc é a boa resposta do algoritmo em relação às bases de dados utilizadas nesta tese para os dados obtidos. As bases utilizadas, conforme mencionadas anteriormente são: uma base de redes Sem Fio Ad Hoc coletadas em ambiente acadêmico [8] e, um conjunto de dados disponibilizado publicamente pelo laboratório Lincoln do MIT denominada KDD 99 [74]. Estes dados mostram que Rede Neural Mapa Auto-Organizável - MAO atingem parcialmente o objetivo do trabalho, pois conseguem classificar parte das classes anômalas, sendo então não utilizada na abordagem definida nesta tese.

4.2 ALGORITMO K-MÉDIAS

O algoritmo K-Médias (*K-Means*) é uma técnica de inteligência computacional que agrupa e classifica dados de acordo com a distância, sendo proposto por Macqueen em 1967 [81]. Segundo

Macqueen [81] o algoritmo K-Médias é proposto a fim de compreender várias aplicações, sendo a de maior destaque aplicações que necessitam separar objetos em grupos, denominados *clusters*. Ainda segundo Macqueen [81] estes *clusters* são formados através da aplicação de técnicas de medidas de distâncias entre os objetos.

O K-Médias realiza o agrupamento por meio de otimização, através da aplicação de uma função objetivo. Esta função objetivo baseia-se em protótipos que têm a finalidade de encontrar n *clusters*, sendo o valor de n determinado pelo usuário. Esta quantidade de *clusters* serão representados pelos seus centróides, ou centros de gravidade [82].

A técnica não-hierárquica de agrupamento esta presente no algoritmo K-Médias, que baseia-se em protótipos que serão responsáveis por criarem um particionamento de um determinado nível dos objetos de dados a serem agrupados. O processo não-hierárquico de agrupamento portanto define um número k de classes e também realiza uma classificação inicial de n objetos em k classes, sendo o valor de k determinado pelo usuário antes ou depois do processo de agrupamento.

A nível de programação e de processamento computacional o algoritmo K-Médias é de fácil programação e econômico, ou seja, não sendo necessário um alto poder computacional para executarem aplicações que o utilizam. O algoritmo K-Médias é capaz de processar grandes volumes de dados, sendo que a complexidade de armazenamento do mesmo é $O((m + K)n)$, onde m é o número de pontos e n é o número de atributos [81].

O agrupamento é realizado por uma técnica particional, a qual utiliza uma função de otimização denominada de função objetivo. O algoritmo K-Médias aplica a função objetivo baseada no cálculo do erro quadrático(*Sum of Square Erros - SSE*). A função objetivo tem o princípio de medir a qualidade de um agrupamento, sendo possível calcular o erro de cada ponto de dados até os centroides mais próximos e depois realizar o somatório total dos erros quadráticos. O erro quadrático é definido conforme Equação 4.13:

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist(C_i, x)^2 \tag{4.13}$$

Em que:

- $dist$ é a distância Euclidiana padrão entre dois objetos no espaço Euclidiano;
- C_i determina o centroide que minimiza a Soma de Erros Quadráticos(SSE) do grupo de índice i, representado pela média do grupo. O C_i é definido pela Equação 4.14

$$C_i = \frac{1}{m_i} \sum_{x \in C_i} x \tag{4.14}$$

Assim, a finalidade do algoritmo K-Médias é obter uma partição que minimiza o erro quadrático para um determinado número k fixo de *clusters* ou grupos. Por exemplo, fornecidos os conjuntos diferentes de grupos que são produzidos pelas aplicações do algoritmo K-Médias, é

recomendado a de menor erro quadrático, pois significa que os centróides deste agrupamento possuem uma representação melhor dos pontos do seu grupo [82].

A Figura 4.12 mostra o fluxograma de funcionamento do algoritmo K-Médias, que é composto de seis passos fundamentais. O primeiro, abrange o valor preliminar dos centróides, ou seja, (C1, C2, ...) representa os centróides iniciais. O segundo a distância dos objetos dos centróides, sendo a distância entre o centróide do *cluster* e todos os objetos calculados. A distância euclidiana é usada e depois a matriz de distância na iteração 0 é calculada. Cada coluna na matriz de distância significa um objeto. A distância da matriz na primeira linha corresponde à distância de cada objeto à segunda linha e o primeiro centróide representa a distância de cada objeto no segundo centróide. No terceiro passo, tem-se o agrupamento de objetos, alocando todos os objetos baseados na menor distância. O quarto passo determina os centróides, identificando os componentes de todos os grupos, sendo o novo centróide de cada conjunto determinado com base nessas novas associações. Por fim, o último passo realiza a comparação do último agrupamento de iteração e essa iteração indica que os grupos não são movidos pelos objetos. Portanto, a definição de *cluster K* significa que se tornou estável e não há mais necessidade de iteração [83].

O algoritmo K-Médias possui algumas desvantagens que precisam ser analisadas:

- Em uma base de dados grande, o algoritmo K-Médias não poderá ser eficiente na geração de soluções de qualidade caso sua inicialização não seja bem sucedida, bem como seus centróides iniciais que são os grupos que ficarem mal posicionados no espaçamento de busca;
- Em relação ao desempenho o algoritmo não garante o resultado global ótimo, pois a qualidade final da solução depende dos conjuntos iniciais de *clusters*, podendo afastá-los do resultado ótimo global;
- A escolha inapropriada do valor de k pode resultar em resultados ruins.

O algoritmo K-Médias portanto converge para uma solução fazendo uso de combinações de funções de proximidade e também de tipos de centroides atingindo um estado em que nenhum ponto, ou objeto de dados, mude de grupo, por consequência não haverá mudança de centroide, conforme apresentado no pseudocódigo abaixo. Em alguns casos poderá ocorrer do algoritmo K-Médias não atingir estes objetivos, sendo necessário atribuir uma condição mais fraca para atingir o estado final, como por exemplo repetir este processo até que apenas 1% dos objetos não mudem de grupo.

```
1 - Selecione K pontos como centróides iniciais
2 - Repita
3 -   Construa K grupos atribuindo cada ponto ao seu centróide mais próximo
4 -   Recalcule o centróide de cada grupo
5 -Até que os centróides não mudem
```

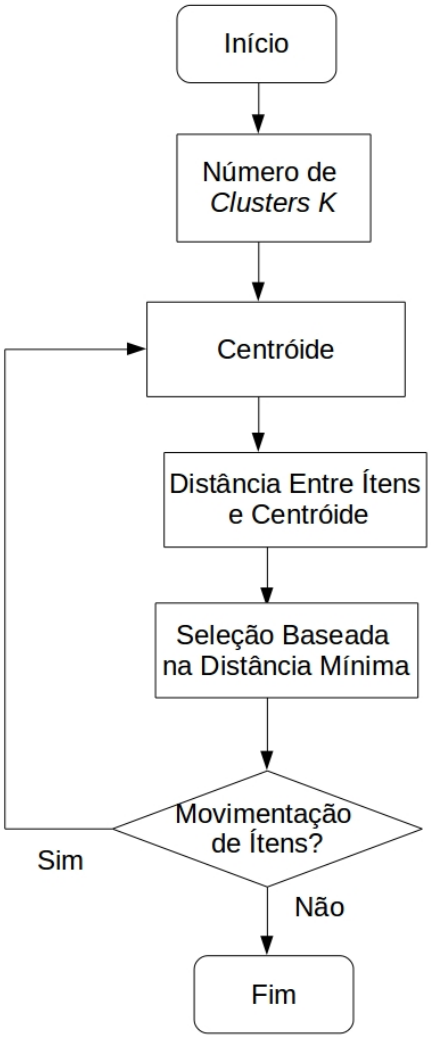


Figura 4.12: Fluxograma de Funcionamento do Algoritmo K-Médias.

K-Médias são algoritmos bem utilizados em aplicações que compartilham padrões comportamentais [81], tendo dificuldades quando se deseja detectar grupos com formas ou tamanhos diferentes, pois não trabalha com grupos chamados de não globulares, de tamanho e densidades diferentes. Em contrapartida o algoritmo K-Médias já se torna bastante eficiente em aplicações em que há a necessidade de múltiplas execuções com frequência.

Algumas aplicações de atuação do algoritmo K-Médias são apresentadas a seguir:

- O algoritmo K-Médias pode ser utilizado para realizar o agrupamento de textos;
- Registrar dados que possuem características semelhantes em um repositório de dados, como um *data mart* por exemplo;
- Utilizado em classificação não-supervisionada, em que são especificadas em quantos *clusters* os objetos de um arquivo devem ser agrupados, respeitando determinados critérios que são adicionados ao próprio algoritmo.

O algoritmo K-Médias é utilizado na proposta desta tese para o agrupamento dos dados oriundos de redes sem fio Ad Hoc, sendo que sua aplicação encontra-se na primeira etapa do sistema proposto, logo após a obtenção dos dados de auditoria da rede Ad Hoc. A definição de sua utilização fundamenta-se em alguns fatores, dentre os quais esta a característica do próprio algoritmo de ser de fácil programação e de baixo consumo de processamento. Também leva-se em consideração o fato deste algoritmo ser eficiente em aplicações com a necessidade de realizar execuções com frequência. Outro aspecto, que favoreceu a sua escolha é a presença deste algoritmo em propostas recentes que envolvam redes *wireless* e/ou redes Ad Hoc [84, 85, 86, 87], sendo normalmente utilizada em uma etapa anterior a classificação.

Outra propriedade que contribuiu para a definição de utilização do algoritmo K-Médias é a boa resposta em relação às bases de dados utilizadas nesta tese para classificação dos dados obtidos. As bases utilizadas, conforme mencionadas anteriormente são: uma base de redes sem fio Ad Hoc coletadas em ambiente acadêmico [8] e, um conjunto de dados disponibilizado publicamente pelo laboratório Lincoln do MIT denominada KDD 99 [74].

4.3 COMENTÁRIOS FINAIS

Neste Capítulo é apresentado o termo Inteligência Computacional e as técnicas de inteligência computacional utilizadas na proposta desta tese, que são : Redes Neurais Artificiais, mais precisamente os algoritmos *MultiLayer Perceptron*, bem como o algoritmo *K-Médias*. A Rede Neural do tipo Mapas Auto Organizáveis - MAO, também é apresentada e avaliada para o sistema proposto.

O próximo Capítulo será apresentado o Sistema de Detecção e Classificação de Intrusão em Redes Ad Hoc, bem como, os resultados e as análises realizadas dos mesmos. O Sistema de Detecção e Classificação proposto é composto por duas etapas que consistem em realizar o agrupamento dos dados da rede para posterior classificação, através dos algoritmos *K-Médias e MultiLayer Perceptron*.

5 PROPOSTA DO SISTEMA DE DETECÇÃO E CLASSIFICAÇÃO DE INTRUSÃO EM REDES AD HOC

Esta tese apresenta uma proposta de Sistema de Detecção e Classificação de Intrusão para redes sem fio Ad Hoc. A proposta faz uso de duas etapas, sendo a primeira destinada a agrupar os dados de auditoria oriundos da rede Ad Hoc, utilizando o algoritmo K-Médias, enquanto que a segunda etapa consiste em realizar a classificação das anomalias previamente conhecidas pela Rede Neural, ou de confirmação de situação normal.

5.1 PROPOSTA COM UTILIZAÇÃO DO ALGORITMO K-MÉDIAS E REDES NEURAIS ARTIFICIAIS

O Sistema de Detecção e Classificação proposto é aplicado localmente em cada componente da rede sem fio Ad Hoc, ou seja, é um sistema de arquitetura *stand-alone* ou autônomo baseado em anomalias, com o objetivo de pela análise dos dados de rede Ad Hoc classificar as anomalias previamente conhecidas pela Rede Neural.

Esta tese é pautada em experimentos computacionais a partir de base de dados existentes, sendo tratados a fundamentação teórica em redes sem fio Ad Hoc, segurança da informação, Sistemas de Detecção de Intrusão e inteligência computacional, mais precisamente em relação às técnicas de aprendizagem supervisionadas e não supervisionadas. A fundamentação teórica é realizada a partir da avaliação de trabalhos relacionados escolhidos de acordo com o número de citações em bases de pesquisas científicas(periódicos e anais de eventos), bem como, sua importância para o objetivo desta tese.

O Sistema de Detecção e Classificação de redes Ad Hoc possui como princípio a utilização de técnicas de inteligência computacional supervisionadas e não supervisionadas. A definição destas técnicas é realizada inicialmente através da seleção das técnicas com melhores desempenhos identificadas nas pesquisas realizadas entre os trabalhos relacionados em periódicos científicos, bem como, em anais de eventos internacionais da área de redes de computadores. Assim, é necessário verificar a eficiência das técnicas de inteligência computacional identificadas através da utilização de dados de redes sem fio Ad Hoc.

O sistema recebe quadros da rede sem fio Ad Hoc denominados de dados de auditoria, ou seja, são informações úteis do tráfego capturado para a identificação de determinadas anomalias previamente definidas. O sistema então realiza o agrupamento destes dados através do algoritmo K-Médias, sendo incluído na base de dados a informação do *cluster* a qual cada quadro pertence. Assim, além das informações relacionadas ao pacote de rede sem fio , tem-se também a informação da classe anômala ou normal, assim como, o *cluster* respectivo. No entanto, realiza-se a clas-

sificação dos dados com a Rede Neural Artificial do tipo *MultiLayer Perceptron*, sendo que para a experimentação deste trabalho possuem 17 neurônios de entrada, 20 neurônios na camada oculta e com cinco classes de saídas(*Normal, EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*).

A saída do sistema gera-se *logs*, sendo um contendo as informações para os dados classificados de maneira correta, mas com classes anômalas. Nestes casos grava-se as informações retiradas do próprio sistema operacional, tais como data, horário, bem como, o endereço de origem e destino, porta e anomalia. Já, para os quadros que não são classificados de maneira correta, gera-se um *log* com todas as informações para posterior análise dos administradores da rede em questão.

O processo de avaliação é realizado através da utilização do Sistema Gerenciador de Banco de Dados *PostgreSQL*, bem como do *software* Weka (*Waikato Environment for Knowledge Analysis*) [88]. Weka consiste em um pacote público na linguagem Java composto por um conjunto de implementações de algoritmos de diversas técnicas de Inteligência Computacioal, tais como: Classificação, Clusterização, Associação, Seleção de Variáveis.

Esta tese propõe um Sistema local de Detecção e Classificação em redes Ad Hoc em duas etapas: a primeira baseada no algoritmo K-Médias, para o agrupamento dos dados, para posterior detecção de comportamentos anômalos, através de Redes Neurais Artificiais.

A primeira etapa do Sistema de Detecção e Classificação proposto fundamenta-se na utilização do algoritmo K-Médias, que é uma técnica de inteligência computacional com o objetivo de separar determinados objetos em grupos, chamados de *clusters*. Estes *clusters* são gerados através da aplicação de técnicas de medidas de distâncias, bem como, técnicas de similaridades entre os objetos [81].

O término desta primeira etapa, apresenta os dados de auditoria previamente classificados e rotulados com o grupo ou *cluster* a qual pertence cada quadro capturado de rede Ad Hoc. Esta informação de qual *cluster* pertence determinado registro é fundamental para a otimização do processo de classificação que será realizado pela Rede Neural, pois fará parte do conjunto de informações que mais fortemente identificam as classes de anomalias aprendidas pela Rede Neural.

A segunda etapa é formada por uma Rede Neural Artificial, com cinco neurônios na camada de saída. Essa Rede Neural é treinada para reconhecer quatro classes distintas de anomalias em redes Ad Hoc, conforme apresentada na Tabela 5.1, além do tráfego normal.

A Rede Neural Artificial desta etapa é um *Multilayer Perceptron - MLP* que será treinada por *backpropagation* [89]. A camada de entrada terá 17 neurônios, como apresentado na Figura 5.1. Na Figura 5.1 o último neurônio de entrada é o *Cluster* oriundo da saída do algoritmo K-Médias, após a realização do agrupamento dos dados de auditoria da rede Ad Hoc e rotulação dos mesmos.

O algoritmo para o Sistema de Detecção e Classificação de Intrusão proposto, de forma simplificada é apresentado na Figura 5.2. Os dados da rede sem fio Ad Hoc são obtidos através da captura do tráfego da rede. Note que a captura deste tráfego é realizada a cada inicialização do

Tabela 5.1: Classes de Anomalias em Redes Ad Hoc

Classe	**Propriedade**	**Descrição**
Normal	Tráfego Normal	Identifica um comportamento normal da rede Ad Hoc
EAPOLStart	Método de autenticação	Carga excessiva de solicitação *EAPOL - Start*, sobrecarregando dispositivos de interconexão dos dispositivos da rede.
BeaconFlood	Identificação da localização do BSS(*Basic Set Service*)	São solicitações do tipo gerenciamento, que tem a finalidade de transmitir milhões de *Beacons* não válidos, resultando na dificuldade que determinado dispositivo da rede terá na identificação de um IBSS(*Independent Basic Service Set*) legítimo [53].
Deauthentication	Desautenticação de dispositivos da Rede	São solicitações do tipo gerenciamento, que são injetados na rede. Os quadros pertencentes a esta anomalia são transmitidos como pedidos imaginários, os quais solicitão a desautenticação de um dispositivo que se encontra autorizado.
RTSFlood	Negação de Serviço	Transmissão em grande escala de pacotes ou *frames RTS* por um curto período de tempo. A inundação de *frames RTS* na rede Ad Hoc proporcionará o congestionamento na reserva do canal, resultando no processo de negação de serviço aos dispositivos da rede [53].

algoritmo, afim de não gerar prejuízo aos recursos dos dispositivos da rede. O algoritmo, então, realiza um pré-processamento dos dados, redefinindo-os com as variáveis pertinentes a serem utilizadas nos próximos passos. Exemplos de informações importantes para arquitetura de redes Ad Hoc incluem:

- Banda disponível;
- Banda utilizada;
- Número de Fluxos;
- Número de Pacotes enviados e/ou recebidos;
- Quantidade de bytes enviados e/ou recebidos;
- Quadros de Associação;
- Taxa de Transmissão.

O próximo passo após realizar o pré-processamento dos dados obtidos da rede sem fio Ad Hoc é a geração dos *clusters* através da execução do algoritmo K-Médias. Para a execução do algoritmo K-Médias define-se a construção de 25 *clusters* e a utilização da função de distância

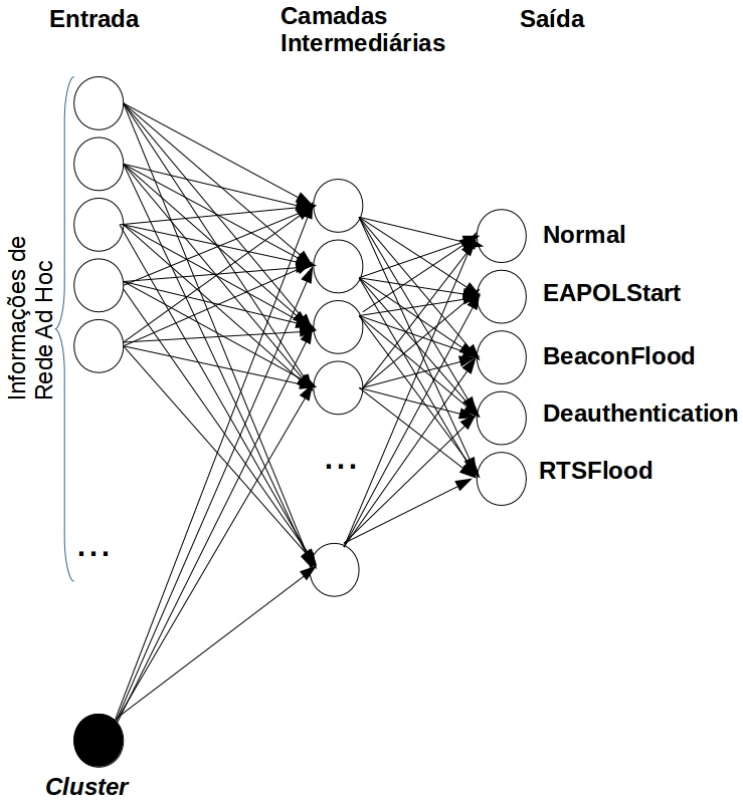


Figura 5.1: Arquitetura da Rede Neural *Multilayer Perceptron*.

Euclidiana para medir a similaridade entre os dados de cada grupo. A definição do número de *clusters* é definida de forma empírica para que se alcance os melhores resultados de classificação.

Após a rotulação dos dados, ou seja, identificação a qual *cluster* cada quadro de tráfego da rede sem fio Ad Hoc pertence, realiza-se o segundo pré-processamento dos dados de forma a incluir nos dados a informação da classe pertencente a cada registro. Os dados podem ser classificados em quatro anomalias(*EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*) e também em comportamento normal.

O terceiro passo do algoritmo proposto é realizar a classificação dos dados, através do treinamento da Rede Neural *Multilayer Perceptron*. Para a execução da Rede Neural *Multilayer Perceptron* é definida uma Rede Neural com 17 neurônios na entrada(16 referentes à rede sem fio Ad Hoc e 1 referente ao *cluster*) e 10 neurônios para a camada oculta, como mostra a Figura 5.1. A Rede Neural *Multilayer Perceptron* de acordo com o sistema proposto é treinado com dados preexistentes, contendo o conhecimento das anomalias definidas, bem como de estado normal. O conjunto de dados contém informações obtidas de quadros capturados da redes Ad Hoc, juntamente com o rótulo de *cluster* pertencente a cada quadro obtido.

Para os dados que são classificados em uma das anomalias predefinidas, implica que uma tentativa de intrusão é identificada, então o sistema proposto gera um arquivo de *log*, contendo as informações dos dados que serão úteis para o administrador do ambiente. Os dados classificados corretamente os *Verdadeiro Negativos e Verdadeiro Positivos*, que representam com-

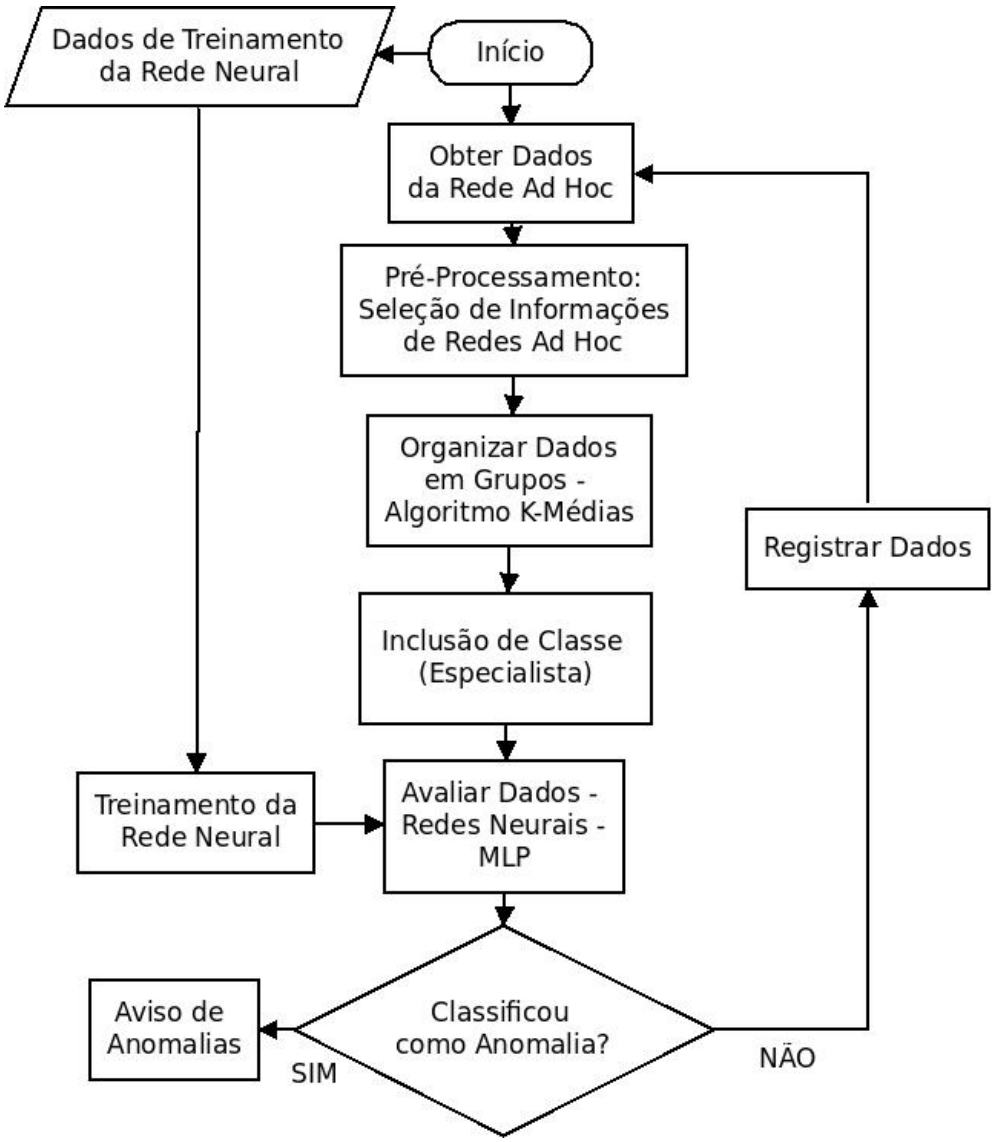


Figura 5.2: Diagrama Geral da Proposta.

portamento normal classificado como "normal"e comportamento anormais classificados em suas respectivas classes(*EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*). Os dados classificados incorretamente são os Falso Positivos e Falso Negativos, que representam eventos "normais"classificados como anômalos e eventos anômalos classificados como "normais"respectivamente. Já os dados que não são classificados corretamente são também registrados em um arquivo de *log* para posteriori análise de especialista. No entanto, estes dados classificados de maneira incorreta, são definidos pela métrica de SDI denominada de *Falso Positivo e Falso Negativo*, que refletem comportamento normal em algumas das classes anômalas(*EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*) e comportamento anômalos na classe "normal"respectivamente.

Para avaliar computacionalmente a o sistema proposto é utilizado o software Weka Data Mining (*Waikato Environment for Knowledge Analysis*) [90] e aplicado a um conjunto de dados de auditoria. O projeto WEKA fornece uma coleção ampla de algoritmos de aprendizado de máquina

e ferramentas de pré-processamento de dados para pesquisadores e profissionais. No entanto, concede que pesquisadores realizem experimentos e comparações rápidas entre diferentes técnicas de aprendizado de máquina. A arquitetura deste software é modular e extensível, permitindo que novos algoritmos de mineração de dados sejam construídos, através da vasta coleção de algoritmos e ferramentas fornecidas. Assim, estender o *framework* é fácil devido a uma API(*Application Programming Interface*) simples, mecanismos de *plug-in* e recursos que automatizam a integração de novos algoritmos de aprendizado com as interfaces gráficas do WEKA. O algoritmo K-Medias e a Rede Neural *Multilayer Perceptron* estão dispostos nas seções de classificadores e clusterização respectivamente.

Este sistema proposto é bastante simples e flexível, sendo possível sua utilização em diversos tipos de arquiteturas de redes de computadores. Para isso, deve-se realizar um pré-processamento dos dados que contenha características importantes da arquitetura de rede utilizada. Em uma rede sem fio Ad Hoc, pode-se utilizar quadros da camada de aplicação, a exemplo dos quadros de solicitação de associação em pontos de acesso. Com isso, será possível identificar anomalias exclusivas dessa arquitetura. Como os dados são obtidos diretamente na rede, não é esperado aumento de *overhead*, independente da arquitetura utilizada.

5.2 RESULTADOS

A proposta desta tese é a detecção e classificação de anomalias em redes sem fio Ad Hoc através da utilização de um sistema em duas etapas. A primeira etapa é responsável pela organização dos dados da rede Ad Hoc em grupos ou *clusters*, enquanto que a segunda etapa objetiva-se em classificar os dados agrupados com as classes pré-definidas(*EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*). Assim, é apresentado nesta seção os resultados obtidos com as simulações realizadas, bem como, as métricas de avaliação utilizadas, juntamente com a análise dos resultados encontrados e algumas comparações com propostas afins.

5.2.1 Base de Dados

A Base de Dados é um importante componente na validação da proposta apresentada, pois contribuirá para a análise dos dados necessários da rede sem fio Ad Hoc para a classificação de anomalias. Também são importantes para avaliar a eficiência dos algoritmos principalmente na definição de dados de característica de *Falso Positivos e Falso Negativos*.

Assim, define-se duas bases de dados a serem utilizadas nas simulações presentes nesta tese. Estas bases são responsáveis por contribuírem para a definição das técnicas de inteligência computacional utilizadas na proposta deste trabalho, como também na própria avaliação do sistema proposto. Para a definição dos algoritmos que compõe a proposta desta tese, estes são validados separadamente utilizando as bases de dados apresentadas a seguir.

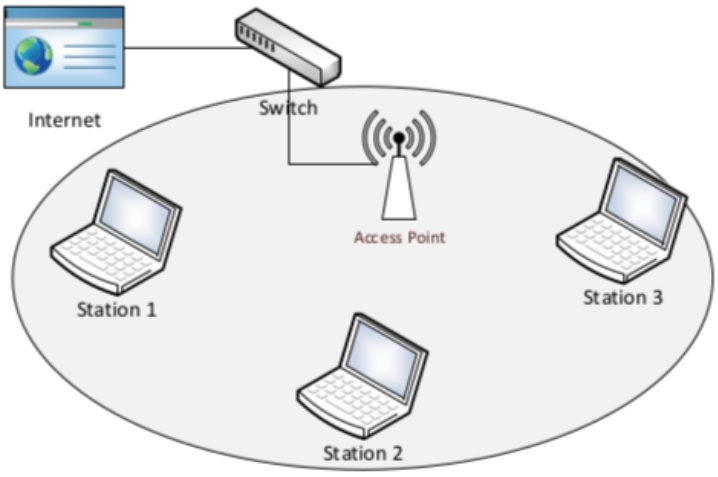


Figura 5.3: Exemplo de Topologia de Ambiente Doméstico [8].

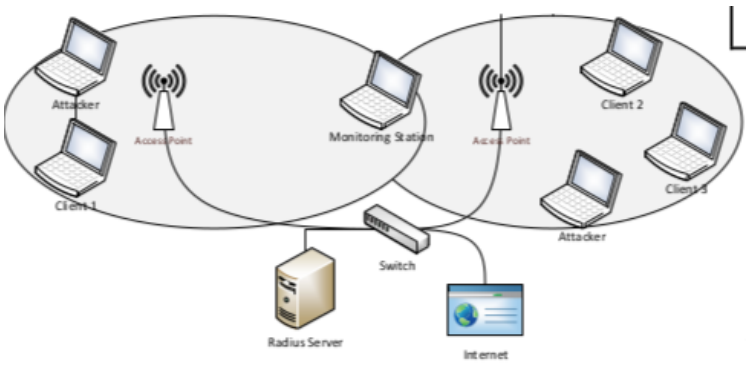


Figura 5.4: Exemplo de Topologia de Ambiente Corporativo [8].

5.2.1.1 Base de Dados de Redes *Wireless* e Ad Hoc

A base de dados de redes *wireless* e Ad Hoc representam uma coleção real de tráfego de rede capturados na arquitetura *Wireless*. Estes dados por sua vez mostram o comportamento de usuários que frequentemente utilizam esta arquitetura de rede para acessar diversas informações, bem como para a utilização da Internet. De acordo com [8], para esta base de dados utilizou-se o tráfego de rede obtido por estudantes e funcionários da instituição em que realizou-se o experimento.

A base de dados obtida fez uso de dois cenários distintos, aumentando as possibilidades de tráfego em redes *wireless*. Os cenários abordados possuem configurações e topologias próprias, sendo um cenário representando um ambiente doméstico típico destas redes, enquanto que o outro cenário apresenta um ambiente mais complexo, corporativo. Os dois ambientes são apresentados pelas Figura 5.3 e Figura 5.4 respectivamente.

O primeiro cenário trata-se de uma simples topologia, sendo comumente utilizado em ambiente doméstico e em pequenas empresas.

De acordo com Ferreira [8], a criação do conjunto de dados é definida utilizando ataques do tipo *Denial of Service - DoS*. Estes tipos de anomalias são escolhidas pois são relativamente simples o seu uso, e até usuários com pouco conhecimento técnico são capazes de executá-los, pois existem diversos softwares prontos que implementam os ataques. Esta categoria de ataque explora vulnerabilidades nos quadros de gerenciamento, causando impacto na disponibilidade dos serviços em redes IEEE 802.11 com criptografia pré-RSN (*Robust Security Network*).

Ainda conforme Ferreira [8], é utilizado o software *Aircrack* [**??**] para geração de ataques do tipo *ChopChop* [91], desautenticação, fragmentação e duração, pela Estação *Station1*. Entretanto, a Estação *Station2* é utilizada para realizar a captura dos dados, com o uso do software *Wireshark* [92]. A estação *Station3* é utilizada para gerar tráfego normal na rede, com aplicações que fazem uso dos protocolos HTTP e HTTPS.

Neste cenário optou-se pelo emprego da captura baseada em espaço amostral, para permitir o uso do conjunto de dados mesmo em ambientes com pouco poder computacional, ainda que o conjunto seja reduzido, existe representatividade com amostras de todos os tipos de tráfego [8].

Para o segundo cenário são utilizados quatro tipos de ataques que são: Desautenticação, *Beacon Flood*, *RTS-Flood* e *EAPOL-Start*. Os ataques de desautenticação e *Beacon Flood* possuem como alvos os quadros de gerenciamento, enquanto que o *RTS-Flood*, possuem além dos quadros de gerenciamento também os quadros de dados como alvos, e o *EAPOL-Start* visa os quadros de controle.

Os ataques de desautenticação ocorrem quando o atacante gera quadros falsos em broadcast, endereço “FF:FF:FF:FF:FF:FF” na rede. A estação que recebe este quadro automaticamente se desconectará da rede. Este processo é então repetido continuamente [54].

Os *beacons* são quadros de sincronismo enviados periodicamente pelos dispositivos da rede com informações sobre a mesma. O seu emprego também possui a função de auxiliar na sincronização na rede, geralmente são transmitidos a cada 100ms. Contudo, no ataque de *Beacon Flood*, é produzido número demasiado de quadros com objetivo de impossibilitar clientes de se associarem a rede.

Os quadros Requisição para Enviar – RTS são enviados pelos dispositivos de rede sem fio Ad Hoc quando existem dados para serem transmitidos, e assim fazer a reserva do canal para comunicação. Todavia, de acordo com Ferreira [8] com objetivo de causar danos na rede, os quadros RTS utilizados nos ataques tiveram o campo duração alterado para um valor elevado, provocando a paralização da mesma.

Os quadros *EAPOL* são utilizados para transportar segmentos de rede oriundos do Protocolo Extensível de Autenticação – EAP sobre uma Rede Local – LAN, com objetivo de prover a comunicação entre um cliente (requisitante) e o ponto de acesso (autenticador). Ataques de *EAPOL-Start* geram excessivas requisições de inicializações de sessões *EAPOL* a dispositivos da rede sem fio Ad Hoc, caracterizando-se como um Ataque de Negação de Serviço – DoS, com objetivo de paralisar o dispositivo.

Esta base de dados é composta por um total de 616047 registros, sendo que cada registro é composto por 16 variáveis que representam características do próprio tráfego de rede *wireless*, que são:

- *Protocol Version* - Indica a versão corrente do protocolo 802.11 utilizado. As estações receptoras usam esse valor para determinar se a versão do protocolo do quadro recebido é suportada;
- *Type* - Determina a função do quadro. Há 3 diferentes tipos de quadro: controle, dados e gerenciamento;
- *Subtype* - Há múltiplos subtipos para cada tipo de quadro. Cada subtipo determina uma função específica desempenhada com o seu tipo de quadro associado;
- *To DS* - Indica se o quadro está indo para o DS;
- *From DS* - Indica se o quadro é oriundo de DS;
- *Retry* - Indica se a informação (dado ou gerenciamento) está ou não sendo retransmitida;
- *Power Management* - Indica se a estação que transmitiu a informação está em *active mode* (modo ativo) ou em *power-save-mode* (modo economia de energia);
- *More Data* - Indica para uma estação operando em *power-save-mode* que o AP tem mais quadros para enviar. Isso é também usado por AP's para indicar que quadros de *broadcast/ multicast* adicionais estarão sendo enviados;
- *WEP* - Indica ou não se está sendo usado no quadro o processo de criptografia e autenticação. Isso pode ser configurado para todos os quadros de dados e gerenciamento que têm o *subtype* configurado para autenticação;
- *Order* - Indica se todos os quadros de dados recebidos devem ser processados em ordem;
- *Duration* - Esse campo é usado para todos os campos de controle, exceto com o *subtype* chamado *Power Save (PS) Poll*, para indicar o tempo restante necessário para receber a próxima transmissão. Quando é usado o subtipo *PS Poll*, esse campo contém a AID (*Associaton Identity*) da estação que está transmitindo. Para a reserva virtual usando-se CTS/RTS esse campo contém o período de tempo que o meio irá ficar ocupado;
- *BSS Indentifier* - BSSID (Identificador de BSS): BSSID unicamente identifica cada BSS. Quando o quadro é vindo de uma estação que opera em modo infra-estrutura BSS, BSSID é o endereço MAC do AP. Quando o quadro é vindo de uma estação que opera em modo ad hoc (IBSS), o BSSID é um número randômico gerado e localmente administrado pela estação que iniciou a transmissão;
- *Destination Address* - DA (Endereço Destino): Indica o endereço MAC do destino final para a recepção do quadro;

- *Source Address* - AS (Endereço Fonte): Indica o endereço MAC da fonte que originou (criou) e transmitiu inicialmente o quadro;
- *Receiver Address* - RA (Endereço do Receptor): Indica o endereço MAC da próxima estação que irá receber o quadro;
- *Transmitter Address* - TA (Endereço do Transmissor): Indica o endereço MAC da estação que transmitiu o quadro na rede sem fio;
- *Sequence Control* - Indica o número para cada fragmento do quadro enviado. O valor inicial é zero e é incrementado para cada fragmento;

Também em cada registro da base de dados é apresentado como última informação a classe a qual pertence determinado registro, classificação esta realizada levando em consideração os dados referente ao tráfego de rede obtido. Desta forma os dados estão classificados em:

- *Normal*: Dados que possuem características de tráfego de redes wireless aceitáveis;
- *EAPOLStart*: Uso do protocolo *Extensible Authentication Protocol*(EAP), cujo o objetivo é realizar um método de autenticação tanto na utilização do protocolo *Wired Equivalent Provacy*(WEP), tanto para o protocolo *Wi-Fi Protected Access*(WPA), em suas versões comerciais para acesso a redes wireless. Esta anomalia se caracteriza por uma carga excessiva de solicitação *EAPOL - Start*, que em um sobrecarregamento do *Access Point*, responsável pela interconexão dos dos dispositivos da rede Wireless;
- *BeaconFlood*: Solicitações do tipo gerenciamento, que têm a finalidade de transmitir milhões de *Beacons* não válidos, resultando na dificuldade que determinado disposito da rede Wireless terá na identificação de um *Access Point* legítimo. Este quadro *Beacons* tem contribuição nos dispositivos na identificação da localização do BSS(*Basic Set Service*) de uma Rede Wireless [53];
- *Deauthentication*: Solicitações do tipo gerenciamento, que são injetados na Rede Wireless. Os quadros pertencentes a esta anomalia são transmitidos como pedidos imaginários, os quais solicitão a desautenticação de um dispositivo que se encontra autorizado na Rede Wireless;
- *RTSFlood*: Denominado *Request-to-Send Flood* é um quadro do tipo controle. Esta anomalia se baseia na transmissão em grande escala de pacotes ou frames RTS por um curto período de tempo. A inundação de frames RTS na Rede Wireless proporcionará o congestionamento na reserva do canal Wireless, resultando no processo de negação de serviço aos nós da Rede Wireless [53].

Um fragmento do Conjunto de Dados é mostrado na Figura 5.5. Os dados são separados por vírgula, sendo que cada linha reflete um quadro do trafego que é obtido da rede sem fio.

```
0,2,8,0,1,0,0,0,1,1,0,633564802256,525400501505,633564802256,266753615,60,Normal
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal
0,2,8,0,1,0,1,0,1,1,0,1326311586,525400501505,1326311586,266753615,60,Normal
0,0,11,0,0,0,1,0,0,0,0,266753600,2264345062,266753600,2264345062,314,Normal
0,0,11,0,0,0,0,0,0,0,0,266753615,918633546526,266753615,918633546526,0,EAPOLStart
0,1,13,0,0,0,0,0,0,0,0,0,143036511504,0,0,Normal
0,0,11,0,0,0,1,0,0,0,0,16193335064,266753615,16193335064,266753615,314,EAPOLStart
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753600,666666666666,266753600,0,Normal
0,1,13,0,0,0,0,0,0,0,0,0,0,266753612,0,0,Normal
0,2,0,0,1,0,0,0,0,0,0,944239466088,266753626,944239466088,266753626,314,Normal
0,0,11,0,0,0,0,0,0,0,0,143036511504,266753615,143036511504,266753615,314,Deauthentication
0,2,8,1,1,0,1,0,0,0,0,243696817522,525400501505,266753626,266753615,44,Normal
0,1,13,0,0,0,0,0,0,0,0,0,0,266753626,0,0,Normal
0,1,11,0,0,0,0,0,0,0,0,0,0,266753615,805662419435.000064,1016,RTSFlood
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,BeaconFlood
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal
0,0,5,0,0,0,1,0,0,0,0,685079219663.000064,266753615,685079219663.000064,266753615,314,Normal
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753626,666666666666,266753626,0,Normal
0,1,13,0,0,0,0,0,0,0,0,266753600,0,0,0,0,Normal
```

Figura 5.5: Fragmento do Conjunto de Dados.

Entretanto, observa-se que a última coluna indica o tipo de anomalia ou comportamento normal, consequentemente, trata-se de uma base rotulada, conforme já mencionado anteriormente. Esta informação é fundamental para a eficiência da Rede Neural *Multilayer Perceptron*, responsável pela classificação dos dados oriundos da rede sem fio Ad Hoc no sistema proposto. A Tabela 5.2 descrimina a característica representada por cada coluna na Figura 5.5.

Tabela 5.2: Descrição do Fragmento de Dados

Coluna	**Descrição**
1	*Protocol Version*
2	*Type*
3	*Subtype*
4	*To DS*
5	*From DS*
6	*Retry*
7	*Power Management*
8	*More Data*
9	*WEP*
10	*Order*
11	*Duration*
12	*BSS Identifier*
13	*Destination Address*
14	*Receiver Address*
15	*Transmitter Address*
16	*Sequence Control*
17	*Classe*

O processo de rotulação dos quadros capturados da rede sem fio Ad Hoc sao realizados manualmente através da análise das variáveis. No entanto, há a possibilidade de construção de ferramentas que a partir de informações existentes possa alimentar a base de dados. Para a proposta desta tese sugere a aplicação do Algoritmo 1, levando em consideração as variáveis que melhor representam as classes. Deve-se salientar que para este algoritmo os quadros já existentes

e rotulados são armazenados em uma matriz *m* x *n*.

Algoritmo 1: ROTULAÇÃO DE QUADROS CAPTURADOS - REDES AD HOC

```
/* A entrada representa informações oriundas do quadro
   capturado de rede Ad Hoc e quadros[m][n] representa os
   quadros já conhecidos e rotulados                        */
```

Entrada: $subtype, morefragment, retry, duration, transmitteraddress, quadros[m][n]$

Saída: Inclusão do campo $classe(EAPOLStart, BeaconFlood, Deauthentication, RTSFloodeNormal)$ para cada quadro capturado

```
1 início
2     m ← número de linhas da base de dados
3     para m de 1 até m faça
4         se (quadro[m][3] = subtype) e (quadro[m][6] = morefragment) e
           (quadro[m][7] = retry) e (quadro[m][12] = duration) e
           (quadro[m][15] = transmitteraddress) então
5             classe ←quadro[m][17]
6         fim
7     fim
8 fim
```

5.2.1.2 KDD 99

Para avaliar a referida proposta, também utiliza-se um conjunto de dados disponibilizados publicamente pelo laboratório Lincoln do MIT denominada KDD 99. A construção desta base foi realizada em 1999 utilizando a captura de tráfego, através da ferramenta *tcpdump* [74]. Também são inseridos dados que simulam certos tipos de ataques a três alvos baseados em sistemas operacionais distintos. A topologia da rede que originou a base de dados é apresentada na Figura 5.6.

Os dados capturados são extraídos de conexões que são representadas por uma linha. Cada linha contém 41 propriedades, dentre as quais tem-se: duração, protocolo, aplicação, *flags*, entre outros. As conexões possuem uma etiqueta que informa sua classificação como normal ou o nome do ataque. Na Figura 5.7 são apresentados exemplos de conexões do KDD 99.

Uma conexão é uma sequência de pacotes TCP com início e término em um intervalo de tempo, com fluxo de dados entre um dispositivo origem e destino, em relação a um determinado protocolo. Esses dados são processados e condensados em quatro conjuntos [93, 2]:

- Características básicas: constituídas por informações dos cabeçalhos dos pacotes sem a análise dos dados transportados. Apresentam informações básicas das conexões e são detalhadas na Tabela 5.3;

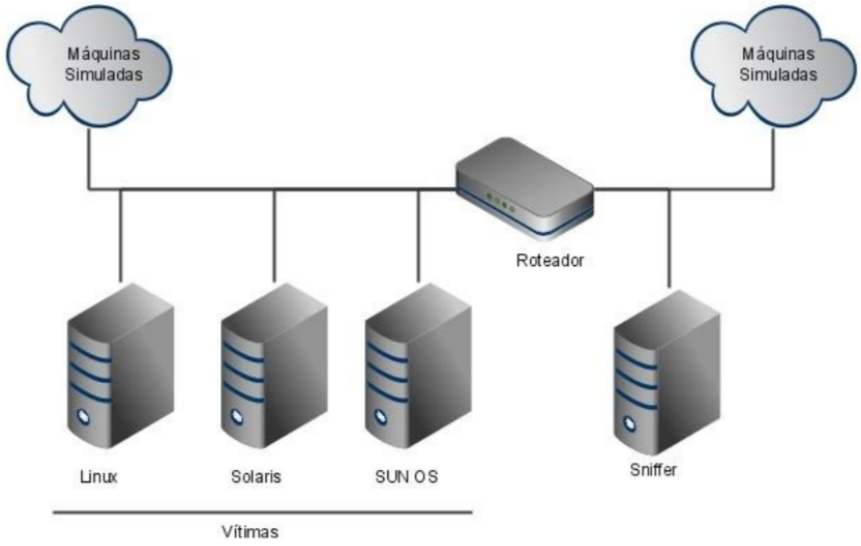


Figura 5.6: Topologia da Rede KDD 99 [2].

5,tcp,smtp,SF,959,337,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,
1.00,0.00,0.00,144,192,0.70,0.02,0.01,0.01,0.00,0.00,normal.

0,tcp,http,SF,54540,8314,0,0,0,2,0,1,1,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.
00,0.00,0.00,118,118,1.00,0.00,0.01,0.00,0.00,0.00,0.02,0.02,back.

0,tcp,http_443,S0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,114,2,1.00,1.00,0.00,0.00,0.0
2,0.06,0.00,255,2,0.01,0,07,0.00,0.00,1.00,1.00,1.00,0.00,neptune.

Figura 5.7: Exemplos de Conexões KDD 99 [2].

- Características sugeridas: são utilizados conhecimentos de especialistas para extrair informações do conteúdo dos pacotes transportados. Isso inclui, por exemplo, o número de tentativas de *login* com erros. São apresentados na Tabela 5.4;
- Características de tráfego com janela de 2s: são características de tráfego calculadas no intervalo de tempo de 2s. Um exemplo é o número de conexões destinadas a um mesmo dispositivo da rede. Na Tabela 5.5 são apresentadas essas características;
- Características de tráfego das últimas cem conexões: essas métricas apresentam o perfil histórico das últimas cem conexões realizadas. Um exemplo dessas métricas é a quantidade de conexões para um mesmo dispositivo de destino, são apresentadas na Tabela 5.6.

A base de dados do KDD 99 apesar de ser antiga tem diversos trabalhos que fazem uso da mesma, a exemplo de [88, 94, 95, 96]. Entretanto, a base de dados do KDD 99 tornou-se referência para análise de propostas de Sistema de Detecção de Anomalias, sendo utilizada em trabalhos recentes, [97], [98], [99] e [100], por ter uma elevada representatividade para essas análises.

Tabela 5.3: Características Básicas do KDD 99

Nome	Descrição	Tipo
duration	Duração da conexão em segundos	Contínuo
protocol_type	Protocolo da camada de transporte (TCP, UDP)	Discreto
service	Protocolo da camada de aplicação (HTTP, Telnet, etc)	Discreto
src_bytes	Quantidade de bytes transmitidos da origem para o destinatário	Contínuo
dst_bytes	Quantidade de bytes transmitidos do destinatário para a origem	Contínuo
flag	Indica o estado da conexão (normal, erro)	Discreto
land	Possui valor 1 se o número de porta origem e de destino são iguais e 0 caso diferentes	Discreto
wrong_fragment	Número de fragmentos errados	Contínuo
urgent	Indica o número de pacotes marcados como urgente	Contínuo

5.2.1.3 Seleção de Atributos

As bases de dados com muitos atributos, demandam alto consumo de processamento computacional, para execução de processos de classificação e reconhecimento de padrões. Este é um problema identificado na utilização da base KDD 99, pois possui um total de 41 atributos. Neste caso, torna-se importante a redução do número de atributos, visto que esta quantidade de característica poderá requerer grande poder computacional quando são avaliados as propostas de Sistemas de Detecção de Intrusão.

A seleção de atributos é importante para projetos de Sistemas de Detecção de Intrusão. Assim, somente as características mais importantes são extraídas das bases de dados. A finalidade é prevenir que características irrelevantes possam causar ruídos no processo de classificação e reconhecimento de padrões [101].

Diferentes propostas têm sido empregadas neste aspecto, porém, o processo para seleção dos melhores atributos de uma base de dados permanece o mesmo, conforme Zuech[101]. A metodologia de seleção de atributos utilizada para a base de dados KDD 99 é apresentada na Figura 5.8. A partir do conjunto original, são gerados subconjuntos, até que seja encontrado um subconjunto ótimo, que é escolhido conforme um determinado critério, no entanto, trata-se de um processo

Tabela 5.4: Características Sugeridas KDD 99

Nome	**Descrição**	**Tipo**
hot	Indica o número de pacotes importantes	Contínuo
num_failed_logins	Número de falhas de *login*	Contínuo
logged_in	O valor 1 indica que o *login* foi efetuado com sucesso e o valor 0 caso contrário	Discreto
num_compromised	Indica o número de conexões "comprometedoras"	Contínuo
root_shell	O valor 1 indica que o shell do root foi obtido e 0 caso contrário	Discreto
su_attempted	O valor 1 indica que houve tentativa de se obter o shell do root e o valor 0 caso contrário	Discreto
num_root	Número de acessos como root	Contínuo
num_file_creations	Indica o número de operações com criação de arquivos	Contínuo
num_shells	Indica o número de shell abertos	Contínuo
num_access_files	Indica o número de operações no controle de acesso de arquivos	Contínuo
num_outbound_cmds	Número de comandos executados numa sessão de FTP	Contínuo
is_hot_login	O valor 1 indica se o *login* pertence a uma lista importante e 0 caso contrário	Discreto
is_guest_login	O valor 1 indica se o *login* foi executado como convidado e 0 caso contrário	Discreto

iterativo. Assim, o subconjunto obtido é menor que o conjunto original, desta forma, o número de características no subconjunto ótimo é menor que a quantidade de características existente no conjunto original.

Um critério de parada que pode ser utilizado é denominado de *Ganho de Informação*. Este critério fundamenta-se na efetividade de um atributo em classificar um conjunto de treinamento. Logo, informa o quanto determinado atributo é bom para classificar o conjunto de dados. Esta medida é realizada através do cálculo da redução da entropia do conjunto quando o atributo é selecionado para classificar o conjunto [101].

Tabela 5.5: Características de Tráfego com Janela de 2s no KDD 99

Nome	**Descrição**	**Tipo**
count	Número de conexões ao mesmo nó nos últimos 2s	Contínuo
serror_rate	Percentual de conexões com erros do tipo SYN para o mesmo host	Contínuo
rerror_rate	Percentual de conexões com erros do tipo REJ para o mesmo host	Contínuo
same_srv_rate	Percentual de conexões ao mesmo serviço para o mesmo host	Contínuo
diff_srv_rate	Percentual de conexões a serviços diferentes para o mesmo host	Contínuo
srv_count	Número de conexões para o mesmo serviço de um mesmo nó	Contínuo
srv_serror_rate	Percentual de conexões com erros do tipo SYN para os mesmos serviços	Contínuo
srv_rerror_rate	Percentual de conexões com erros do tipo REJ para os mesmos serviços	Contínuo
srv_diff_host_rate	Percentual de conexões a diferentes nós para os mesmos serviços	Contínuo

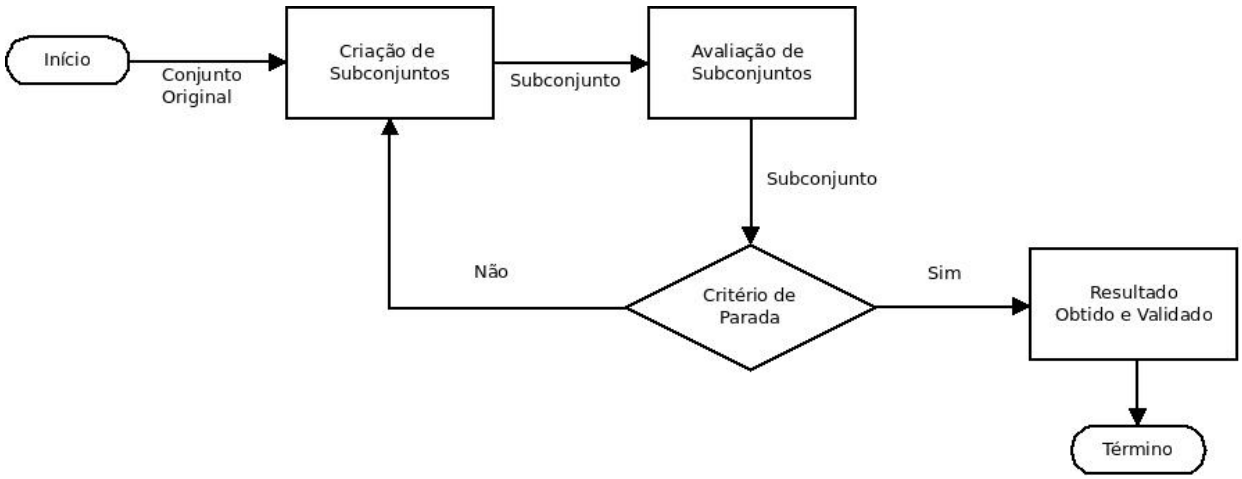


Figura 5.8: Processo de Seleção de Atributos.

A base do algoritmo de Ganho de Informação está no cálculo da entropia de uma distribuição de probabilidade. A entropia pode ser definida como a forma de medir o quão dispersa se encontra a distribuição de probabilidade. Esta pode ser determinada através da equação 5.1 [102]:

Tabela 5.6: Características de Tráfego Utilizando as Últimas 100 Conexões no KDD 99

Nome	**Descrição**	**Tipo**
dst_host_count	Número de conexões para o mesmo nó destino	Contínuo
dst_host_srv_count	Número de conexões para o mesmo nó destino e o mesmo serviço	Contínuo
dst_host_same_srv_rate	Percentual de conexões para o mesmo nó destino e o mesmo serviço	Contínuo
dst_host_diff_srv_rate	Percentual de conexões de serviços diferentes para o mesmo host atual	Contínuo
dst_host_same_src_port_rate	Percentual de conexões para o nó atual e mesma porta	Contínuo
dst_host_srv_diff_host_rate	Percentual de conexões para o mesmo serviço com diferentes nós de origem	Contínuo
dst_host_serror_rate	Percentual de conexões ao nó atual com um erro S0	Contínuo
dst_host_srv_serror_rate	Percentual de conexões para o nó atual e serviço especificado com um erro S0	Contínuo
dst_host_rerror_rate	Percentual de conexões para o nó atual que tiveram um erro RST	Contínuo
dst_host_srv_rerror_rate	Percentual de conexões para o nó atual e serviço especificado que tiveram um erro RST	Contínuo

$$H(X) = -\sum_{i=1}^{n} P(X)_i log_2 P(X)_i \tag{5.1}$$

onde:

$H(X)$: entropia da distribuição de probabilidade X;

$P(X)$: a probabilidade de acontecimento de cada um dos valores X;

n: o número de valores distintos que X toma.

Desta maneira, quando a entropia é baixa irá traduzir-se numa maior facilidade de prever qual o valor que, neste caso, X irá tomar. Entretanto, quando a entropia é alta então há uma distribuição uniforme onde todos os valores de X têm a mesma probabilidade de ocorrência, diminuindo a

probabilidade de prever qual o valor que X irá tomar.

Assim, a segunda distribuição de probabilidade é uma particularização da primeira, isto é, a primeira distribuição de probabilidade condicionada a um dos seus valores. Então tem-se [102]:

$$H(X|Y) = -\sum_{i=1}^{n} P(Y)_i \sum_{j=1}^{m} P(X|Y)_j log_2 P(X|Y)_j \tag{5.2}$$

onde:

Y: é um dos valores que X pode tomar.

O *Ganho de Informação* pode então ser definido como a diferença entre entropias de duas distribuições [102].

$$IG(X|Y) = -\sum_{i=1}^{n} P(X)_i log_2 P(X)_i + \sum_{i=1}^{n} P(Y)_i \sum_{j=1}^{m} P(X|Y)_j log_2 P(X|Y)_j \tag{5.3}$$

onde:

$IG(X|Y)$: é o Ganho de Informação que o valor Y possui para a determinação de X.

De acordo com pesquisas realizadas, a quantidade de características que podem ser desprezadas da base de dados KDD 99 é relevante, sendo que das 41 características deste conjunto de dados, 29 podem ser descartadas, por isso apenas 12 são utilizadas pelo Sistema de Detecção de Intrusão, obtendo bons resultados [101]. O estudo sobre o *Ganho de Informação* das características utilizadas para a base de dados KDD 99 neste trabalho é mostrado na Tabela 5.7.

5.2.1.4 Avaliação da Base de Dados

A avaliação das bases de dados é realizada através do emprego de técnicas de classificação habitualmente utilizadas em Sistemas de Detecção de Intrusão, sendo estes métodos com aprendizagem supervisionada e não supervisionada. A avaliação da eficiência destas técnicas de classificação em relação às bases de dados é realizada através de medidas de erro apuradas durante a fase de treinamento e por meio da medida do percentual de classificação na avaliação, além da comparação da medida do coeficiente *Kappa*, descritos a seguir.

O Erro Médio Absoluto -EMA é definido como a média da diferença entre os valores calculados e os valores reais. O cálculo do Erro Médio Absoluto é mostrado na Equação 5.4, onde *n* representa o número de termos, x_i é o valor real do *i-ésimo* termo, e x'_i é o valor calculado do *i-ésimo* termo. Os valores mais próximos de zero indicam que ocorreram melhores classificações [8].

Tabela 5.7: Métrica *Ganho de Informação* - KDD 99

Ganho de Informação	**Atributo**
1.779485	src_bytes
1.605052	count
1.234414	srv_count
1.140227	dst_host_diff_srv_rate
1.139148	dst_host_same_src_port_rate
1.128007	dst_host_srv_count
1.091182	flag
1.075448	dst_host_same_srv_rate
1.006437	diff_srv_rate
0.966289	same_srv_rate
0.945934	protocol_type
0.939919	dst_host_serror_rate
0.93563	serror_rate
0.842841	dst_bytes
0.817373	logged_in
0.694611	dst_host_srv_diff_host_rate
0.675393	dst_host_count
0.659198	dst_host_srv_serror_rate
0.644539	srv_serror_rate
0.530555	dst_host_rerror_rate
0.517101	rerror_rate
0.378545	dst_host_srv_rerror_rate
0.366194	srv_rerror_rate
0.298306	srv_diff_host_rate

$$EMA = \frac{\sum_{x=1}^{n} |x_i - x_i'|}{n} \tag{5.4}$$

A Média da Raiz Quadrada do Erro – MRQE é a média do quadrado do erro, a fórmula para o cálculo é exibida na Equação 5.5, onde *n* representa o número de termos, x_i o valor real do *i-ésimo* termo, e x_i' o valor calculado do *i-ésimo* termo. O valor do Erro Médio Absoluto mínimo, as vezes não pode indicar uma variação mínima, assim, o emprego em conjunto com a Média da Raí Quadrada do Erro torna-se importante [8].

$$MRQE = \sqrt{\frac{\sum_{x=1}^{n} |x_i - x_i'|^2}{n}} \tag{5.5}$$

Estes indicadores fornecem uma estimativa simples sobre a eficácia das técnicas de inteligência computacional empregadas na avaliação das bases de dados. Seu uso é introdutório, assim, é necessário utilizar tanto o Erro Médio Absoluto, tanto a Média da Raiz Quadrada do Erro para melhorar a comparação.

O coeficiente *Kappa* é uma métrica de concordância induzida, primeiramente usada entre

observadores de psicologia. Esta métrica mede o grau de aceitação ou de respostas concordantes entre diversos juízes. É empregado utilizando a proporção entre a concordância observada (*Po*) e a concordância devida ao acaso (*Pa*), sendo o cálculo realizado pela Equação 5.6 [8].

$$k = \frac{Po - Pa}{1 - Pa} \tag{5.6}$$

O valor unitário indica que a classificação é correta, enquanto que o valor do coeficiente nulo, indica que a classificação ocorreu ao mero acaso, portanto valores próximos de 1 confirmam que são empregados os melhores classificadores. Este coeficiente tem sido utilizado em propostas para implementação de Sistema de Detecção de Intrusão [103, 104, 105, 8].

A avaliação das bases de dados é executada com implementação das técnicas de inteligência computacional definidas neste trabalho que são: Rede Neural *Multilayer Perceptron*, Rede Neural Mapa Auto-Organizável - MAO e Algoritmo K-Médias. A avaliação é realizada inicialmente de maneira individual para cada uma destas técnicas de inteligência computacional, para posterior avaliação do sistema proposto nesta tese. Estas técnicas são definidas levando em consideração sua utilização em propostas de Sistemas de Detecção de Intrusão em redes sem fio Ad Hoc, bem como, pela eficiência demonstrada pelas métricas Erro Médio Absoluto, Média da Raiz Quadrada do Erro, Coeficiente *Kappa*, Falsos Positivos e Falsos Negativos.

5.2.2 Experimentos

Os experimentos desta tese são organizados em duas etapas distintas, sendo a primeira destinada a avaliar de maneira indivual técnicas de inteligência computacional em relação às bases de dados escolhidas e verificando a eficiência das mesmas, de acordo com as métricas apontadas anteriormente. Deve-se salientar que os resultados obtidos são fundamentais para a escolha das técnicas que compuseram o Sistema de Detecção e Classificação de Intrusão proposto neste trabalho.

A segunda etapa de experimentos fundamenta-se em avaliar o Sistema de Detecção e Classificação de Intrusão através da utilização da base de dados de ambiente de redes *wireless* com tráfego de redes Ad Hoc [8], bem como dados oriundos de redes Ad Hoc doméstica, conforme cenários apresentados anteriormente.

Para as duas etapas de experimentos citadas a avaliação das técnicas de inteligência computacional adotadas é realizada a partir das métricas de avaliação Erro Médio Absoluto, Média da Raiz Quadrada do Erro, Coeficiente *Kappa*, Falsos Positivos, Falsos Negativos e taxa de classificação corretas.

Para todos os experimentos os testes são realizados atendendo o processo de avaliação denominado de validação cruzada ou *Cross-Validation*. Validação cruzada é uma técnica que avalia a capacidade de generalização de uma realidade, através de um conjunto de dados. Esta técnica é utilizada em modelos que há a necessidade de predição, em que busca-se o melhor desempenho

Figura 5.9: Estrutura Método *K-Fold* [9].

de um modelo para novos conjuntos de dados [9].

A técnica de validação cruzada caracteriza-se no particionamento da base de dados em subconjuntos que são reciprocamente exclusivos, sendo utilizados alguns destes subconjuntos para a estimação das propriedades de treinamento, enquanto que os demais subconjuntos são utilizados para validação da base de dados.

O método de particionamento adotado pela técnica de validação cruzada nesta tese é o método *k-fold*. Este método tem a propriedade de dividir a base de dados em *k* subconjuntos mutuamente exclusivos do mesmo tamanho e, assim, um subconjunto é utilizado para teste e os *k-1* restantes são utilizados para estimação dos parâmetros, calculando a precisão dos dados. Este processo é realizado *k* vezes alternando de forma circular o subconjunto de teste. A Figura 5.9 apresenta a estrutura do método *k-fold* [9].

Após as *k* iterações o desempenho final é calculado através da média de acerto de todas *k* classificações da validação *k-fold*.

5.2.2.1 Rede Neural *MultiLayer Perceptron*

A Rede Neural *MultiLayer Perceptron - MLP*, pertence a técnica de inteligência computacional Redes Neurais Artificiais com aprendizado supervisionado, ou seja, necessitam da presença de um especialista no processo de aprendizagem. Esta Rede Neural exige ao menos uma camada escondida com função de ativação não-linear [6].

Para a realização da validação e verificação da Rede Neural *Multilayer Perceptron* utilizam-se as bases de dados mencionadas anteriormente, sendo uma com dados de redes *wireless* e Ad Hoc [8] e, a outra com dados do KDD 99. Para ambas as bases de dados determina-se uma parte dos dados para treinamento e outra para a validação da referida técnica, através da técnica de particionamento validação cruzada, com o método *k-fold*, sendo utilizado 10% dos dados para teste de maneira aleatória.

Para a base de dados de redes *wireless* e Ad Hoc, a Figura 5.10 mostra o percentual de classificação para a técnica de validação cruzada em relação a Rede Neural *Multilayer Perceptron*.

A Tabela 5.8 apresenta a discriminação dos dados em cada classe, enquanto que a Tabela 5.9 mostra a taxa de classificação total dos dados. A matriz de confusão da base de dados aplicada à Rede Neural *Multilayer Perceptron* é apresentada na Figura 5.11. A matriz de confusão apresenta os valores classificados corretamente para a classe indicada, bem como os valores classificados de forma incorreta. Por exemplo a primeira linha e coluna "a"na figura 5.11 representam os valores

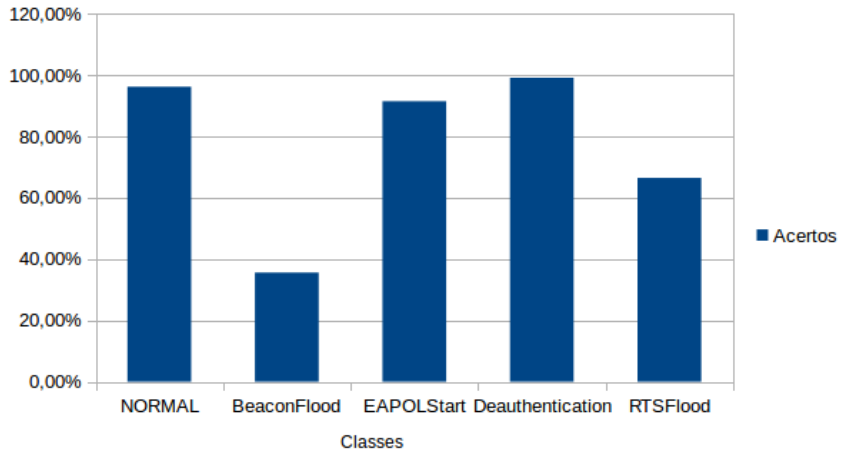


Figura 5.10: Classificação Base de Dados *Wireless* e Ad Hoc - Validação Cruzada.

```
=== Confusion Matrix ===

      a      b      c      d      e   <-- classified as
 523469  14952    468   5260     37 |      a = Normal
  17668   9756      0      0      0 |      b = BeaconFlood
   2221      0  25679    171      0 |      c = EAPOLStart
     39      0    103  16075      0 |      d = Deauthentication
     50      0      0      0     99 |      e = RTSFlood
```

Figura 5.11: Matriz Confusão - Base de Dados *Wireless* e Ad Hoc - MLP.

classificados corretamente, enquanto que as demais colunas representam os dados classificados incorretamente.

Tabela 5.8: Dados Base de Dados *Wireless* e Ad Hoc - Validação Cruzada

Classe	**Acertos**	**Erros**
Normal	523469	20716
BeaconFlood	9756	17668
EAPOLStart	25679	2392
Deauthentication	16075	142
RTSFlood	99	50

Tabela 5.9: Classificação Total - Base de Dados *Wireless* e Ad Hoc - Validação Cruzada

Acertos	**Erros**
93,35%	6,65%

A Tabela 5.10 mostra os valores obtidos para as métricas de avaliação da Rede Neural *Multilayer Perceptron* que são: Erro Médio Absoluto, Média da Raiz Quadrada do Erro, Coeficiente *Kappa*, Falsos Positivos e Falsos Negativos.

Entretanto, a Figura 5.12 apresenta o percentual de classificação dos dados do KDD 99 para a técnica de validação cruzada em relação a Rede Neural *Multilayer Perceptron*.

Tabela 5.10: Métricas de Avaliação *Multilayer Perceptron*

Erro Médio Absoluto	**Média da Raiz Quadrada do Erro**	**Coeficiente *Kappa***	**Falso Positivo**	**Falso Negativo**
0.0333	0.1351	0.6923	3,8%	27,8%

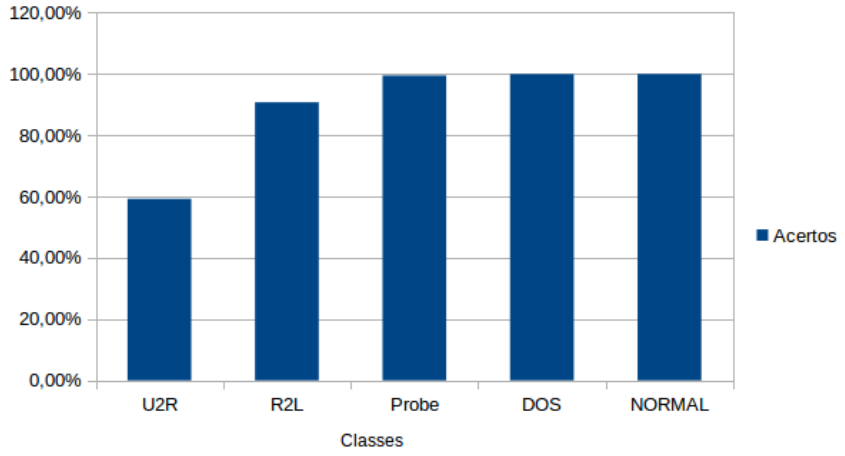


Figura 5.12: Classificação do KDD 99 - Validação Cruzada.

Todavia, a Tabela 5.11 apresenta a discriminação dos dados em cada classe e a Tabela 5.12 mostra a taxa de classificação total dos dados KDD 99.

Tabela 5.11: Dados KDD 99 - Validação Cruzada

Classe	**Acertos**	**Erros**
U2R	48	33
R2L	1282	131
Probe	52375	288
DoS	154088	102
Normal	100920	50

Tabela 5.12: Classificação Total - Base de Dados KDD 99 - Validação Cruzada

Acertos	**Erros**
99,80%	0,20%

A Tabela 5.13 mostra os valores obtidos para as métricas de avaliação da Rede Neural *Multilayer Perceptron* para a base de dados KDD 99 que são: Erro Médio Absoluto, Média da Raiz Quadrada do Erro, Coeficiente *Kappa*, Falsos Positivos.

Nota-se que a Rede Neural *Multilayer Perceptron* possui uma taxa de classificação aceitável para a base de dados *Wireless* e Ad Hoc, em torno de 93%, significando que consegue classificar de maneira correta 93% dos dados após o treinamento da Rede Neural. Também ressalta-se que esta técnica possui erros médios baixos e coeficiente *Kappa* em 0.6923, ou seja, havendo baixo

Tabela 5.13: Métricas de Avaliação *Multilayer Perceptron*

Erro Médio Absoluto	**Média da Raiz Quadrada do Erro**	**Coeficiente *Kappa***	**Falso Positivo**
0.0003	0.0131	0.9973	0,04%

erro durante a fase de treinamento representando um bom classificador. Assim, afirma-se que esta técnica de inteligência computacional é viável para a utilização em sistemas de classificação envolvendo redes *wireless* e Ad Hoc.

Entretanto, os resultados obtidos na validação da base de dados KDD 99 reforça a eficiência da Rede Neural *Multilayer Perceptron*, sendo classificados corretamente aproximadamente 99% dos dados. Pode-se observar erros médios baixos e coeficiente *Kappa* aceitáveis, ou seja, havendo pouco erro durante a fase de treinamento e validação da mesma, bem como a representação de um ótimo classificador para os dados, pois apresenta coeficiente *Kappa* próximo de um.

5.2.2.2 Mapas Auto-Organizáveis - MAO

Mapas Auto-Organizáveis(*Self-organized map*), também conhecido por Mapas de *Kohonen* são um tipo de Rede Neural Artificial que possuem como princípio fundamental o procedimento competitivo de aprendizado entre as unidades da rede. Este tipo de Redes Neurais Artificiais dispõe como principais aplicações a descoberta de agrupamentos de dados, visto que o aprendizado é não supervisionado.

O objetivo dos Mapas de *Kohonen* é construir sistemas que se organizem internamente através da distribuição dos dados de entrada, sem a presença de um especialista. Neste sentido Mapas Auto-Organizáveis(MAO) apresentarão em sua saída a formação de aglomerados, denominado de *clusters*, que apresentam uma resposta máxima a um determinado estímulo. Assim as Redes Neurais que representam os Mapas de *Kohonen* exibem as seguintes características:

- São compostos por uma única camada contendo entrada e saída;
- A entrada da Rede de *Kohonen* corresponde a um vetor no espaço *d-dimensional* em $\mathbb{R}^d$;
- Cada neurônio de saída possui um vetor no espaço *d-dimensional* em $\mathbb{R}^d$, o qual esta associado ao vetor de entrada;
- Os vetores de saída estão interconectados através da relação de vizinhança de acordo com a estrutura do mapa.

Para a realização da validação e verificação da Rede Neural Mapas Auto-Organizáveis(MAO) utilizam-se as bases de dados mencionadas anteriormente, sendo uma com dados de redes *wireless* e Ad Hoc [8] e, a outra com dados do KDD 99. Para ambas as bases de dados determina-se uma

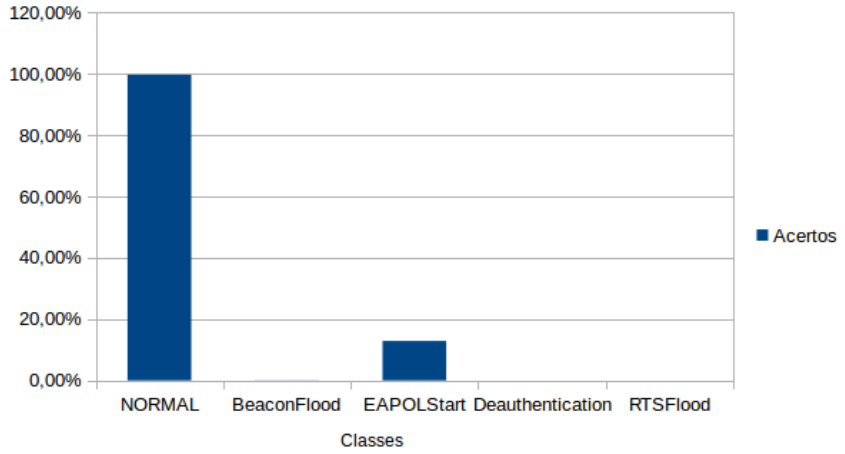


Figura 5.13: Classificação Base de Dados de Redes *Wireless* e Ad Hoc - Validação Cruzada.

parte dos dados para treinamento e outra para a validação da referida Rede Neural, através da técnica de particionamento validação cruzada, com o método *k-fold*, sendo utilizado 10% dos dados para teste de maneira aleatória.

Para a base de dados de redes *wireless* e Ad Hoc, a Figura 5.13 apresenta o percentual de classificação em relação as classes determinadas para a técnica de validação cruzada.

Contudo, a Tabela 5.14 mostra a discriminação dos dados em cada classe e a Tabela 5.15 exibe a taxa de classificação total dos dados. A matriz de confusão da base de dados aplicada à Rede Neural MAO é apresentada na Figura 5.14.

Tabela 5.14: Dados Base de Dados *Wireless* e Ad Hoc - Validação Cruzada

Classe	**Acertos**	**Erros**
Normal	542557	1629
BeaconFlood	15	27409
EAPOLStart	3606	24465
Deauthentication	0	16217
RTSFlood	0	149

Tabela 5.15: Classificação Total - Base de Dados *Wireless* e Ad Hoc - Validação Cruzada

Acertos	**Erros**
88,66%	11,34%

A Tabela 5.16 mostra os valores obtidos para as métricas de avaliação da Rede Neural Mapas Auto-Organizáveis, em relação a base de dados *Wireless* e Ad Hoc, que são: Erro Médio Absoluto, Média da Raiz Quadrada do Erro, Coeficiente *Kappa*, Falso Positivos e Falso Negativos.

A Figura 5.15 mostra o percentual de classificação dos dados do KDD 99 para a técnica de validação cruzada em relação à Rede Neural Mapas Auto-Organizáveis.

A Tabela 5.17 apresenta a discriminação dos dados em cada classe e a Tabela 5.18 mostra a

```
=== Confusion Matrix ===

     a     b     c     d     e   <-- classified as
 54239     0   188     0     0 |     a = Normal
  2714     0     0     0     0 |     b = BeaconFlood
  2196     0   638     0     0 |     c = EAPOLStart
  1611     0     4     0     0 |     d = Deauthentication
     0     0    14     0     0 |     e = RTSFlood
```

Figura 5.14: Matriz Confusão - Base de Dados *Wireless* e Ad Hoc - MAO.

Tabela 5.16: Métricas de Avaliação Mapas Auto-Organizáveis - MAO

Erro Médio Absoluto	**Média da Raiz Quadrada do Erro**	**Coeficiente *Kappa***	**Falso Positivo**	**Falso Negativo**
0.0723	0.1903	0.1467	0,34%	90,85%

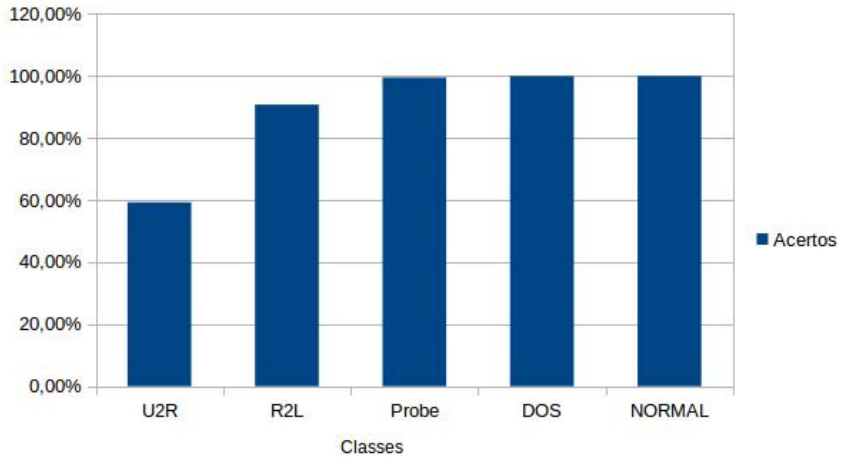


Figura 5.15: Classificação do KDD 99 - Validação Cruzada.

taxa de classificação total dos dados KDD 99.

Tabela 5.17: Dados KDD 99 - Validação Cruzada - MAO

Classe	**Acertos**	**Erros**
U2R	48	33
R2L	1282	131
Probe	52375	288
DoS	154088	102
Normal	100920	50

Tabela 5.18: Classificação Total - Base de Dados KDD 99 - Validação Cruzada - MAO

Acertos	**Erros**
99,80%	0,20%

A Tabela 5.19 mostra os valores obtidos para as métricas de avaliação da Rede Neural Mapas Auto-Organizáveis para a base de dados KDD 99 que são: Erro Médio Absoluto, Média da Raiz Quadrada do Erro, Coeficiente *Kappa*, Falsos Positivos.

Tabela 5.19: Métricas de Avaliação Mapas Auto-Organizáveis

Erro Médio Absoluto	**Média da Raiz Quadrada do Erro**	**Coeficiente *Kappa***	**Falso Positivo**
0.0003	0.0131	0.9973	0,04%

Nota-se que a Rede Neural Mapas Auto-Organizáveis possui uma taxa de classificação abaixo da Rede Neural *Multilayer Perceptron* com 89%, significando que consegue classificar de maneira correta 89% dos dados após o treinamento da Rede Neural, sendo que esta classificação é realizada pelo algoritmo LVQ (*Learning Vector Quantization*). Também ressalta-se que esta técnica possui erros médios altos e coeficiente *Kappa* baixos, ou seja, havendo muito erro durante a fase de treinamento e validação da Rede Neural MAO e que não representa um bom classificador para os dados dispostos. Assim, afirma-se que esta técnica de inteligência computacional não atende o Sistema de Detecção e Classificação proposto nesta tese, pois consegue classificar apenas duas classes de anomalias e também por possuir taxa de classificação inferior às demais técnicas analisadas. Entretanto, pode ser analisado em outras arquiteturas de redes de computadores, como por exemplo em redes com infraestrutura, justificado pelos resultados obtidos para a base de dados KDD 99.

5.2.2.3 Algoritmo K-Médias

O algoritmo K-Médias é uma técnica que utiliza o agrupamento de dados em *K* grupos, também denominado de *K-Médias clustering*. Este algoritmo possui como objetivo encontrar a melhor divisão de *P* dados em *K* grupos C_i, onde $i = 1, 2, 3, \ldots, K$. A distância total entre os dados de determinado grupo e o seu respectivo centro é minimizada.

O funcionamento do algoritmo K-Médias segue as etapas descritas a seguir:

- O algoritmo K-Médias primeiramente atribui de maneira aleatoria os dados *P* em *K* grupos;
- O algoritmo realiza o cálculo das médias dos vetores de cada grupo;
- Nesta etapa cada ponto representado por um dado *P* é deslocado para o seu respectivo grupo, o qual corresponde ao vetor médio mais próximo;
- O algoritmo realiza novamente o cálculo das médias dos vetores e também realiza a distribuição dos dados em cada grupo;
- Este processo de realocação de dados a novos grupos, em que os vetores médios são os mais próximos é realizado até que todos os dados estejam em seus grupos.

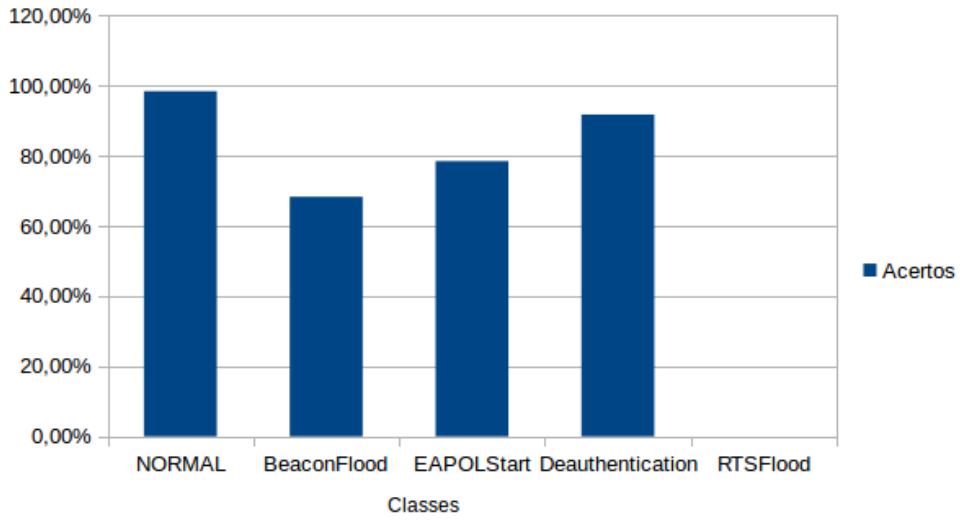


Figura 5.16: Classificação - Algoritmo K-Médias.

Para a realização da validação e verificação do algoritmo K-Médias faz-se uso das bases de dados mencionadas anteriormente, sendo uma com dados de redes *wireless* e Ad Hoc [8] e, a outra com dados do KDD 99. Para ambas as bases de dados determina-se uma parte dos dados para treinamento e outra para a validação do referido algoritmo, através da técnica de particionamento validação cruzada, com o método *k-fold*, sendo utilizado 10% dos dados para teste de maneira aleatória.

A validação do algoritmo K-Médias utilizando a base de dados de redes *wireless* e Ad Hoc [8] emprega um total de 500 iterações e 25 *clusters* ou grupos. A rotulação das classes de anomalias para cada *cluster* é apresentado na Tabela 5.20.

Tabela 5.20: Rotulação dos *Clusters*

	Normal	**EAPOLStart**	**BeaconFlood**	**Deauthentication**	**RTSFlood**
Quantidade de *Clusters*	15	3	6	1	Nenhum

Também realiza-se a porcentagem de acertos e erros para cada classe pré-definidas(*Normal, EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*), ou seja os dados considerados como acertos são aqueles que estão no grupo o qual possui rotulação igual a sua respectiva classe, enquanto que os dados considerados erros são aqueles os quais estão em um grupo o qual possui rotulação diferente a sua respectiva classe. Estes dados são apresentados na Figura 5.16. A Tabela 5.21 apresenta a discriminação dos dados em cada classe.

A Tabela 5.22 mostra a classificação total da base de dados de redes *wireless* e Ad Hoc [8] para a validação do algoritmo *K-Médias*.

A Tabela 5.23 mostra os valores obtidos para as métricas de avaliação do algoritmo K-Médias que são: Tempo de Costrução do Modelo, Falsos Positivos e Falsos Negativos.

Tabela 5.21: Dados Redes *Wireless* e Ad Hoc - K-Médias

Classe	**Acertos**	**Erros**
Normal	53551	876
BeaconFlood	1854	860
EAPOLStart	2224	610
Deauthentication	1482	133
RTSFlood	0	14

Tabela 5.22: Taxa de Percentual de Acertos e Erros - K-Médias - Base de Dalos Redes *Wireless* e Ad Hoc

Acertos	**Erros**
95,95%	4,05%

Tabela 5.23: Métricas de Avaliação Algoritmo K-Médias

Tempo de Costrução do Modelo	**Falso Positivo**	**Falso Negativo**
61,78 segundos	1,61%	46,16%

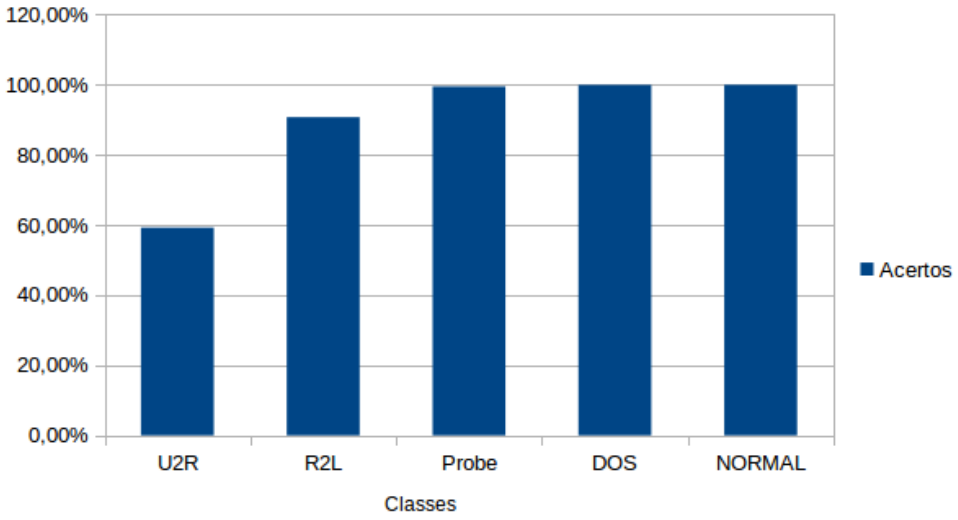


Figura 5.17: Agrupamento dos Dados KDD 99 - K-Médias.

A Figura 5.17 mostra o percentual de classificação dos dados do KDD 99 para a validação do algoritmo K-Médias. Os dados considerados como acertos são aqueles que estão no grupo o qual possui rotulação igual a sua respectiva classe, enquanto que os dados considerados erros são aqueles os quais estão em um grupo o qual possui rotulação diferente a sua respectiva classe.

A rotulação das classes de anomalias para cada *cluster* é apresentado na Tabela 5.24. A Tabela 5.25 apresenta a discriminação dos dados em cada classe.

No entanto, a Tabela 5.26 mostra os valores obtidos para as métricas de avaliação do algoritmo K-Médias que são: Tempo de Costrução do Modelo, Falsos Positivos e Falsos Negativos.

Tabela 5.24: Rotulação dos *Clusters* - KDD 99

	Cluster 0	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Classe	Probe	Normal	DoS	Normal	DoS

Tabela 5.25: Dados KDD9 - K-Médias

Classe	**Acertos**	**Erros**
U2R	48	33
R2L	1282	131
Probe	52375	288
DoS	154088	102
Normal	100920	50

Tabela 5.26: Métricas de Avaliação Algoritmo K-Médias

Tempo de Costrução do Modelo	**Falso Positivo**	**Falso Negativo**
43,20 segundos	0,04%	0,26%

O algorimto K-Médias demonstra através das avaliações nas bases de dados que consegue realizar o agrupamento dos dados em tempos consideravelmente aceitáveis para aplicações de redes *wireless* e Ad Hoc, bem como para Redes de Computadores com infraestrutura. Em relação as métricas de avaliação propostas para este algortimo, tem-se em geral um tempo abaixo de 1 minuto para agrupamento dos dados dispostos(Redes de Computadores Sem Fio e Estruturada), sendo considerável útil para a proposta do Sistema de Detecção e Classificação proposto neste tese.

Outra observância da avaliação do algoritmo K-Médias é a mensuração da taxa de Falso Negativo e Falso Positivo para ambas as bases de dados. Neste processo define-se para cada *cluster* um rótulo, representado por uma classe de anomalia identificada nas bases de dados ou "normal"para comportamento normal dos dados. Este rótulo é definido pela classe que possui mais registros dentro do *cluster*. Diante disto, obtêm-se a taxa de Falso Positivo e Negativo que fundamenta-se na definição dos rótulos, sendo que Falso Positivo representa os dados com rótulo "normal"e encontram-se em *cluster(s)* rotulado(s) com alguma anomalia. Enquanto que Falso Negativo é entendido como dados que tenham rótulos com alguma anomalia e encontram-se em *cluster(s)* rotulado(s) como "normal".

O algoritmo apresenta valores baixos para dados de Redes de Computadores com infraestrutura tanto para a taxa de Falso Positivo, tanto para Falso Negativo, demostrando ser útil em Sistemas de Detecção de Intrusão para estes tipos de redes, conforme dados apresentados na Tabela 5.26 e na Figura 5.17. Entretanto, para os dados de redes *wireless* e Ad Hoc o algoritmo K-Médias apresenta uma taxa de 1,61% para Falso Positivo e 46,16% para Falso Negativo, demonstrando ser útil para um processo de classificação dos dados devido a aproximação de dados com características idênticas.

```
0,2,8,0,1,0,0,0,1,1,0,633564802256,525400501505,633564802256,266753615,60,Normal
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal
0,2,8,0,1,0,1,0,1,1,0,1326311586,525400501505,1326311586,266753615,60,Normal
0,0,11,0,0,0,1,0,0,0,0,266753600,2264345062,266753600,2264345062,314,Normal
0,0,11,0,0,0,0,0,0,0,0,266753615,918633546526,266753615,918633546526,0,EAPOLStart
0,1,13,0,0,0,0,0,0,0,0,0,0,143036511504,0,0,Normal
0,0,11,0,0,0,1,0,0,0,0,16193335064,266753615,16193335064,266753615,314,EAPOLStart
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753600,666666666666,266753600,0,Normal
0,1,13,0,0,0,0,0,0,0,0,0,0,266753612,0,0,Normal
0,2,0,0,1,0,0,0,0,0,0,944239466088,266753626,944239466088,266753626,314,Normal
0,0,11,0,0,0,0,0,0,0,0,143036511504,266753615,143036511504,266753615,314,Deauthentication
0,2,8,1,1,0,1,0,0,0,0,243696817522,525400501505,266753626,266753615,44,Normal
0,1,13,0,0,0,0,0,0,0,0,0,0,266753626,0,0,Normal
0,1,11,0,0,0,0,0,0,0,0,0,0,266753615,805662419435.000064,1016,RTSFlood
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,BeaconFlood
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal
0,0,5,0,0,0,1,0,0,0,0,685079219663.000064,266753615,685079219663.000064,266753615,314,Normal
0,0,8,0,0,0,0,0,0,0,0,666666666666,266753626,666666666666,266753626,0,Normal
0,1,13,0,0,0,0,0,0,0,0,266753600,0,0,0,0,Normal
```

Figura 5.18: Fragmento de Dados - K-Médias.

5.2.2.4 Classificação em Etapas

Este experimento visa avaliar o sistema proposto neste trabalho, que fundamenta-se no processo de detecção e classificação de dados de redes sem fio Ad Hoc. A proposta é pautada através de duas etapas, sendo que a primeira etapa do Sistema de Detecção e Classificação proposto tem como princípio a utilização do algoritmo K-Médias, que é uma técnica de inteligência computacional com o objetivo de separar determinados objetos em grupos, chamados de *clusters*. Estes *clusters* são gerados através da aplicação de técnicas de medidas de distâncias, bem como, técnicas de similaridades entre os objetos [81].

A segunda etapa é formada por uma Rede Neural Artificial do tipo *Multilayer Perceptron*, com cinco neurônios na camada de saída. Essa Rede Neural é treinada para reconhecer quatro classes distintas de anomalias em redes Ad Hoc(*EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*), além do tráfego normal.

Assim, define-se neste experimento a utilização da base de dados de redes *wireless* e Ad Hoc [8], pois apresenta dados reais e mais adequados para a proposta do trabalho. Também define-se o algoritmo *K-Médias* para a realização do agrupamento dos dados presentes na base de dados designada para a validação desta proposta. No entanto a Rede Neural *Multilayer Perceptron - MLP* será utilizada na segunda etapa do sistema proposto, que terá a responsabilidade de classificar os dados. Para a realização da avaliação do sistema faz-se uso da técnica de particionamento de dados denominada validação cruzada ou *cross-validation*, com o método *k-fold*, sendo utilizado 10% dos dados para teste de maneira aleatória.

Após a obtenção dos dados de auditoria e o seu referido pré-processamento, realiza-se em sua primeira etapa o agrupamento dos dados, sendo definido um total de no máximo 500 iterações e 10 *clusters* ou grupos para o algoritmo K-Médias. A Figura 5.18 mostra o fragmento do conjunto de dados utilizado pelo algoritmo K-Médias. Os dados são separados por vírgula, sendo que cada linha reflete um quadro que é obtido da rede. Entretanto, observa-se que a última coluna indica o tipo de anomalia ou comportamento normal, consequentemente, trata-se de uma base rotulada, conforme já mencionado anteriormente.

```
0,0,2,8,0,1,0,0,0,1,1,0,633564802256,525400501505,633564802256,266753615,60,Normal,cluster9
1,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal,cluster8
2,0,2,8,0,1,0,1,0,1,1,0,1326311586,525400501505,1326311586,266753615,60,Normal,cluster3
3,0,0,11,0,0,0,1,0,0,0,0,266753600,2264345062,266753600,2264345062,314,Normal,cluster0
4,0,0,11,0,0,0,0,0,0,0,0,266753615,918633546526,266753615,918633546526,0,EAPOLStart,cluster2
5,0,1,13,0,0,0,0,0,0,0,0,0,143036511504,0,0,Normal,cluster7
6,0,0,11,0,0,0,1,0,0,0,0,16193335064,266753615,16193335064,266753615,314,EAPOLStart,cluster0
7,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753600,666666666666,266753600,0,Normal,cluster8
8,0,1,13,0,0,0,0,0,0,0,0,0,266753612,0,0,Normal,cluster7
9,0,2,0,0,1,0,0,0,0,0,0,944239466088,266753626,944239466088,266753626,314,Normal,cluster9
10,0,0,11,0,0,0,0,0,0,0,0,143036511504,266753615,143036511504,266753615,314,Deauthentication,cluster7
11,0,2,8,1,1,0,1,0,0,0,0,243696817522,525400501505,266753626,266753615,44,Normal,cluster1
12,0,1,13,0,0,0,0,0,0,0,0,0,266753626,0,0,Normal,cluster7
13,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal,cluster8
14,0,0,5,0,0,0,1,0,0,0,0,685079219663.0001,266753615,685079219663.0001,266753615,314,Normal,cluster5
15,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753626,666666666666,266753626,0,Normal,cluster8
16,0,1,13,0,0,0,0,0,0,0,0,0,266753600,0,0,0,0,Normal,cluster7
17,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753612,666666666666,266753612,0,Normal,cluster8
18,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753612,666666666666,266753612,0,Normal,cluster8
19,0,1,13,0,0,0,0,0,0,0,0,0,266753600,0,0,Normal,cluster7
20,0,0,11,0,0,0,0,0,0,0,0,266753615,945454496612.9999,266753615,945454496612.9999,0,EAPOLStart,cluster2
21,0,1,13,0,0,0,0,0,0,0,0,0,266753615,0,0,Normal,cluster7
22,0,2,8,0,1,0,1,0,0,1,0,507574272171,525400501505,507574272171,266753615,60,Normal,cluster9
23,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753626,666666666666,266753626,0,Normal,cluster8
24,0,1,13,0,0,0,0,0,0,0,0,0,0,42633671632,0,0,Normal,cluster7
25,0,1,13,0,0,0,0,0,0,0,0,0,103031527,0,0,Normal,cluster7
26,0,1,13,0,0,0,0,0,0,0,0,0,741223315534,0,0,Normal,cluster7
27,0,1,13,0,0,0,0,0,0,0,0,0,266753615,0,0,Normal,cluster7
28,0,2,0,0,1,0,0,0,0,1,0,10055000062,525400501505,10055000062,266753612,0,Normal,cluster3
29,0,2,8,0,1,0,0,0,0,1,0,681334361412,525400501505,681334361412,266753615,117,Normal,cluster9
30,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal,cluster8
31,0,0,8,0,0,0,0,0,0,0,0,666666666666,266753615,666666666666,266753615,0,Normal,cluster8
32,0,1,13,0,0,0,0,0,0,0,0,0,266753612,0,0,Normal,cluster7
33,0,2,8,1,1,0,1,0,0,1,0,10055000002,266753600,266753612,266753615,44,Normal,cluster1
34,0,1,13,0,0,0,0,0,0,0,0,0,266753612,0,0,Normal,cluster7
35,0,0,1,0,0,0,0,0,0,0,0,143036511504,266753615,143036511504,266753615,314,Normal,cluster6
```

Figura 5.19: Fragmento de Dados de *Cluster*.

A rotulação das classes de anomalias para cada *cluster* é apresentado na Tabela 5.27.

Tabela 5.27: Rotulação de *Clusters*

	Normal	**EAPOLStart**	**BeaconFlood**	**Deauthentication**	**RTSFlood**
Quantidade de *Clusters*	15	3	6	1	Nenhum

O algoritmo K-Médias por sua vez, organiza os dados em 10 *clusters* contendo um tempo de construção do modelo em 14,82 segundos com 19 iterações. Assim, incluí-se na base de dados a informação gerada pelo algoritmo K-Médias, que é o *Cluster* determinado para cada quadro da base de dados, conforme é apresentado no fragmento de dados visualizado na Figura 5.19.

Logo após o término da primeira etapa, passa-se a condução da segunda etapa através da utilização da Rede Neural *Multilayer Perceptron - MLP* com o princípio fundamental de classificar os dados, oriundos da saída da primeira etapa, ou seja, após a execução do algoritmo K-Médias. Esta Rede Neural é composta por uma entrada com 17 neurônios, sendo 16 neurônios com informações da rede e um neurônio com a informação do *cluster*, com uma camada oculta com 20 neurônios e, com cinco classes de saída. A Figura 5.20 mostra a arquitetura da Rede Neural deste experimento.

Para a avaliação da Rede Neural *Multilayer Perceptron* na segunda etapa do sistema proposto, tem-se a Figura 5.21 que mostra o percentual de classificação para a técnica de particionamento, validação cruzada, em relação aos dados resultantes da primeira camada.

A Tabela 5.28 apresenta a discriminação dos dados em cada classe, enquanto que a Tabela 5.29 mostra a taxa de classificação total dos dados. A matriz de confusão da base de dados aplicada à

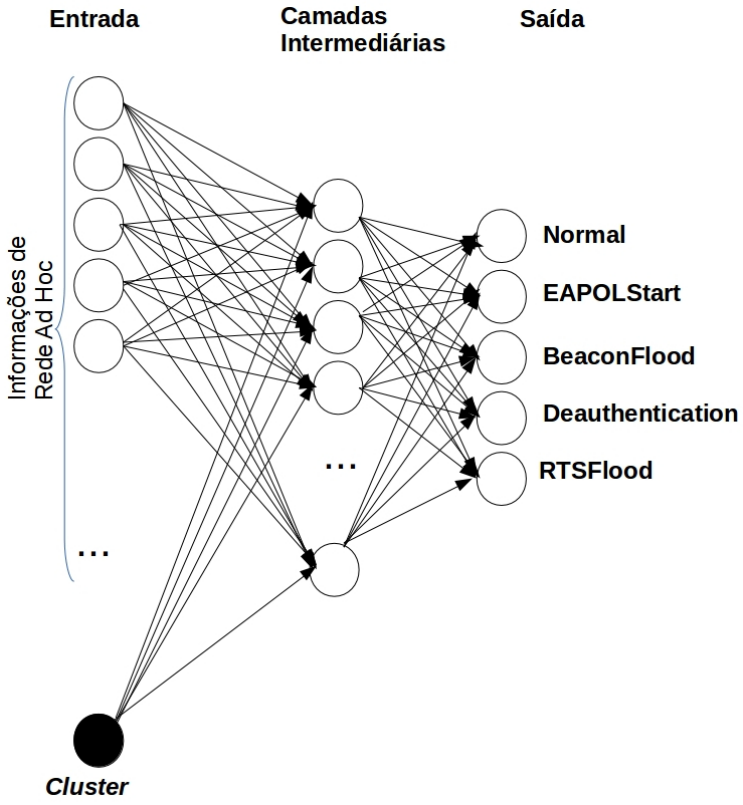


Figura 5.20: Arquitetura da Rede Neural *Multilayer Perceptron*.

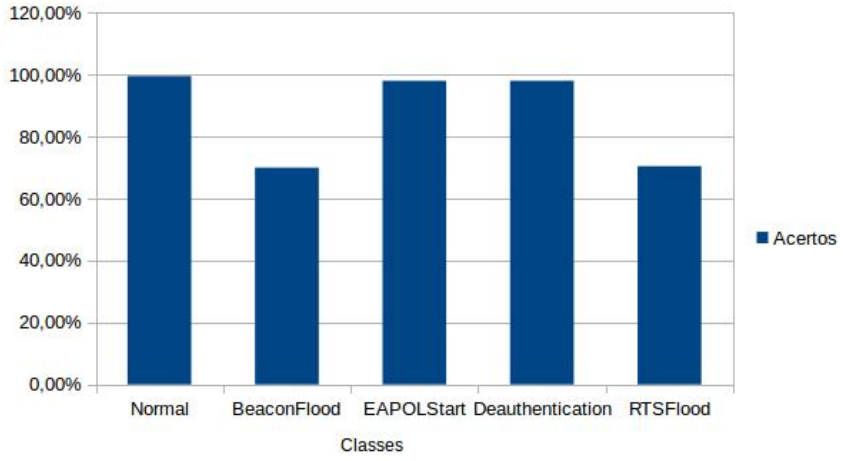


Figura 5.21: Classificação Final - Sistema Proposto.

Rede Neural *Multilayer Perceptron* é apresentada na Figura 5.22.

A Tabela 5.30 mostra os valores obtidos para as métricas de avaliação da Rede Neural *Multilayer Perceptron* que são: Erro Médio Absoluto, Média da Raiz Quadrada do Erro, Coeficiente *Kappa*, Falsos Positivos e Falsos Negativos.

Percebe-se, no entanto que a segunda etapa, através da utilização da Rede Neural *Multilayer Perceptron* possui uma taxa de classificação aceitável de 98%, significando que o Sistema de

Tabela 5.28: Dados Base de Dados *Wireless* e Ad Hoc

Classe	**Acertos**	**Erros**
Normal	541466	2720
BeaconFlood	19205	8229
EAPOLStart	27511	560
Deauthentication	15893	324
RTSFlood	105	44
Total de Dados	616057	

Tabela 5.29: Classificação Total - Sistema Proposto

Acertos	**Erros**
98,07%	1,93%

```
=== Confusion Matrix ===

      a      b      c      d      e   <-- classified as
 541466     30    194   2496      0 |      a = Normal
   8229  19205      0      0      0 |      b = BeaconFlood
    538      0  27511     22      0 |      c = EAPOLStart
     78      0    246  15893      0 |      d = Deauthentication
     44      0      0      0    105|      e = RTSFlood
```

Figura 5.22: Matriz Confusão - Sistema Proposto.

Tabela 5.30: Métricas de Avaliação *Multilayer Perceptron* - Sistema Proposto

Erro Médio Absoluto	**Média da Raiz Quadrada do Erro**	**Coeficiente *Kappa***	**Falso Positivo**	**Falso Negativo**
0.0316	0.127	0.8491	0,50%	12,36%

Detecção e Classificação proposto consegue classificar de maneira correta 98% dos dados após o treinamento da Rede Neural, com dados oriundos da primeira etapa do sistema proposto. Também ressalta-se que esta técnica possui erros médios baixos e coeficiente *Kappa* aceitáveis, ou seja, havendo pouco erro durante a fase de treinamento e validação do algoritmo e que representa um bom classificador. Assim, afirma-se que o Sistema de Detecção e Classificação proposto atende aos objetivos desta tese, que fundamenta-se em classificar de maneira aceitável dados de redes Ad Hoc.

A Tabela 5.31 apresentado um resumo dos resultados encontrados para as etapas definidas no experimento deste trabalho.

Tabela 5.31: Resumo de Experimento

Técnicas	**Base de Dados**	**Erro Médio Absoluto**	**Média da Raíz Quadrada do Erro**	**Coeficiente *Kappa***	**Falso Positivo**	**Falso Negativo**	**Tempo de Construção do Modelo**
Rede Neural *Multilayer Perceptron*	*Wireless* e Ad Hoc	0.0333	0.1351	0.6923	3,8%	27,8%	—-
Rede Neural *Multilayer Perceptron*	KDD 99	0.0003	0.0131	0.9973	0,04%	5%	—-
Rede Neural - MAO	*Wireless* e Ad Hoc	0.0723	0.1903	0.1467	0,34%	90,85%	—-
Rede Neural - MAO	KDD 99	0.0003	0.0131	0.9973	0,04%	5%	—-
Algoritmo K-Médias	*Wireless* e Ad Hoc	—-	—-	—-	1,61%	46,16%	61,78 segundos
Algoritmo K-Médias	KDD 99	—-	—-	—-	0,04%	0,26%	43,20 segundos
Sistema Proposto	*Wireless* e Ad Hoc	0.0316	0.127	0.8491	0,50%	12,36%	61,00 segundos

5.3 DISCUSSÃO DOS RESULTADOS

Os resultados obtidos nas etapas definidas para os experimentos apontam para resultados interessantes na aplicação de técnicas de classificação, também denominada de reconhecimento de padrões, para dados de ambientes de redes sem fio. Para a validação das técnicas de classificação adotadas são utilizados a base de dados KDD 99 e a base de dados apresentada por [8]. As Tabelas 5.9, 5.15, 5.22 mostram a porcentagem de acertos e erros em relação a base de dados de redes *wireless* e Ad Hoc [8] para a Rede Neural *Multilayer Perceptron* e *Mapas Auto-Organizáveis* com o método de aprendizagem por quantização vetorial(LVQ), além do algoritmo de aprendizagem não supervisionado K-Médias respectivamente. Enquanto que as Tabelas 5.12, 5.18, 5.25 apresentam a porcentagem de acertos e erros em relação a base de dados KDD 99 para a Rede Neural *MultiLayer Perceptron* e *Mapas Auto-Organizáveis*, bem como para o algoritmo *K-Médias* respectivamente.

A técnica de inteligência computacional que melhor atende as necessidades de identificação das classes determinadas na base de dados deste trabalho é a Rede Neural *MultiLayer Perceptron* com a utilização de treinamento *backpropagation*. Esta técnica é avaliada individualmente, sendo identificado de forma correta 93,35% dos registros presentes na base de dados de *Wireless* e Ad Hoc, sendo utilizado a técnica de particionamentodos dados, validação cruzada. Entretanto, para a base de dados KDD 99 observa-se a identificação de forma correta dos dados em 99,80% dos dados para a Rede Neural *MultiLayer Perceptron*, levando em consideração as técnicas de validação adotadas.

Em relação a Rede Neural *Mapas Auto-Organizáveis* e o algoritmo K-Médias pode-se verificar que atendem parcialmente o problema de identificação das classes adotadas na base de dados, tanto para a base de dados KDD 99, tanto para a base de dados *Wireless* e Ad Hoc [8]. Para a base de dados *Wireless* e Ad Hoc são classificados corretamente 88,66% dos registros, sendo identificadas apenas as classes *Normal e EAPOLStart*, inviabilizando a sua utilização para o ambiente o qual os dados são coletados. Já para a base de dados KDD 99 há a classificação correta em 95,95% dos registros, sendo identificadas as classes *Normal, U2R, R2L, Probr, DOS*. Apesar da Rede Neural *MultiLayer Perceptron* ser recomendado como o algoritmo que mais atende as necessidade de identificação das classes determinadas, o algoritmo K-Médias poderá alcançar resultados satisfatórios levando em consideração a organização dos dados e quantização dos mesmos.

A eficiência da Rede Neural *MultiLayer Perceptron* também pode ser identificada através da observância do erro médio de cada um deles e na taxa de correteza. Isto significa dizer que para a base de dados KDD 99 tem-se uma saída correta da rede neural em torno de 99,80%, enquanto que para a base de redes *wireless* e Ad Hoc tem-se em torno de 93,35% de saídas corretas da rede neural.

A aplicação dos testes na avaliação do sistema proposto apontam para resultados interessantes, sendo primeiramente organizados os dados em grupos, para depois serem classificados. Para a avaliação do agrupamento dos dados utiliza-se a base de dados de redes *wireless* e Ad Hoc [8] e o algoritmo *K-Médias*, que organiza os mesmos em 25 *clusters* em 19 iterações do algoritmo em 14,82 segundos. A rotulação dos *clusters* é verificada na Tabela 5.27. Para a classificação dos dados já agrupados utiliza-se a Rede Neural *MultiLayer Perceptron*, que possui taxa de acerto de 98,07% dos dados possuindo um erro médio absoluto em torno de 0,0316%, uma média da raiz quadrada do erro de 0,127% e coeficiente *Kappa* em 0,8491. A Tabela 5.29, exibe a taxa de acertos e erros do sistema proposto, enquanto que a Figura 5.21 aponta a classificação em relação as classes abordadas na base de dados.

Diante da apresentação destes dados, bem como, nos conhecimentos adquiridos de Sistemas de Detecção de Intrusão e das técnicas de inteligência computacional adotadas, afirma-se que o sistema proposto se torna viável para o processo de classificação de dados de redes *wireless* e redes Ad Hoc. Deve-se salientar que os dispositivos móveis utilizados nestas avaliações tiveram um bom comportameto, sendo que a eficiência de consumo de energia ainda é um problema para aplicações em redes sem fio Ad Hoc. Assim, deve-se afirmar que o Sistema de Detecção e Classificação proposto não garante a segurança total de ambientes de redes Ad Hoc, sendo necessário a utilização de demais tecnologias em conjunto para prover um ambiente estável, tanto em redes *wireless* e Ad Hoc, tanto em redes dotadas de infraestrutura.

O sistema proposto é competente para o processo de classificação de dados de redes sem fio Ad Hoc, através da utilização de estrutura em etapas, sendo utilizado o algoritmo K-Médias e a Rede Neural *Multilayer Perceptron*, pois permitem a redução de falso positivos e falso negativos, quando comparado ao uso destas técnicas de maneira isolada para o ambiente de redes Ad Hoc

com a base de dados *Wireless* e Ad Hoc. A Tabela 5.32 aponta os valores destas métricas de validação.

Tabela 5.32: Comparação de Falsos Negativos e Positivos

Técnicas	**Falso Positivo**	**Falso Negativo**	**Acertos**
Rede Neural *Multilayer Perceptron*	3,8%	27,8%	93,35%
Algoritmo K-Médias	1,61%	46,16%	95,95%
Sistema Proposto	0,5%	12,36%	98,07%

Existem diversas propostas para construção de Sistemas de Detecção de Intrusão, com abordagens diferentes. A fim de realizar uma comparação entre resultados obtidos, algumas propostas são selecionadas e apresentadas na Tabela 5.33. Para a montagem dessa tabela, são utilizados os valores das taxas de acerto informados diretamente pelos autores, através das publicações de seus trabalhos. Mesmo com abordagens distintas, todas as propostas utilizaram a base de dados de Redes De Computadores. Ainda pela Tabela 5.33, é possível perceber que a proposta aqui apresentada proporcionou uma taxa de acerto aceitável, demonstrando que trata-se de uma abordagem viável para detecção e classificação de dados em ambientes de redes sem fio Ad Hoc.

Tabela 5.33: Comparação entre Trabalhos

Proposta	**Acertos**
SDI Wavelet [21]	99%
SDI Híbrido [20]	98%
Nossa Proposta	98,07%
SDI Thatachi [106]	96%
SDI SVM [16]	96%
MLH-IDS [15]	95%
SDI PCA [27]	91%

O sistema proposto possibilita a implementação de propostas, em que o princípio fundamental seja o agrupamento de dados que tenham alguma similaridade para uma posterior classificação, sendo necessário obter um prévio conhecimento da saída do relacionamento destes dados selecionados. Os dados a serem analisados por propostas derivadas deste sistema proposto, podem ser caracterizados como texto, aúdio, vídeo, etc.

5.4 COMENTÁRIOS FINAIS

Neste Capítulo é apresentada uma proposta para um Sistema de Detecção e Classificação em etapas para classificação de anomalias em redes sem fio Ad Hoc. Essa proposta indica o uso de duas técnicas: Algoritmo Não-Supervisionado K-Médias e Redes Neurais Artificiais *Multilayer Perceptron*. A primeira é responsável pela organização do tráfego da rede em grupos, enquanto

que a segunda realiza sua classificação baseada em um conhecimento prévio de quadros oriundos de redes Ad Hoc rotulados com as seguintes classes anômalas: *EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*.

Nesse capítulo também são apresentados experimentos realizados para a validação da Rede Neural *MultiLayer Perceptron* e MAO, bem como para o algoritmo *K-Médias*. Também é apresentado experimento de validação para o Sistema de Detecção e Classificação proposto envolvendo o algoritmo *K-Médias* e a Rede Neural *MultiLayer Perceptron*.

A abordagem em etapas permite compartilhar as melhores características de cada método. Além disso, o uso em conjunto permite a redução de falso positivos e falso negativos, se comparado com a utilização isolada das estratégias de Inteligência Computacional adotadas. Os resultados obtidos permitem concluir que a abordagem proposta alcança taxas de classificação maiores que as apresentadas quando aplicadas as técnicas de inteligência computacional de maneira individual, se tornado muito promissora, conforme apresenta a Tabela 5.32.

6 CONCLUSÃO

O ambiente de redes *wireless*, mais precisamente as arquiteturas de redes sem fio Ad Hoc, bem como Redes de Sensores Sem Fio possuem em sua característica uma dinamicidade em relação à composição dos integrantes da rede, ou seja, para estes tipos de Redes Sem Fio existem a entrada e saída constante de seus integrantes. Outra característica importante dos dispositivos pertencentes a estas Redes é o consumo da energia dos mesmos, sendo necessário o seu carregamento periodicamente. Estas características trazem a necessidade de um gerenciamento eficiente destes ambientes com o objetivo de ter um maior controle dos mesmos. No entanto, tornam-se bastante vulneráveis a tentativas de abordarem anomalias presentes no sistema como um todo, sendo algumas destas anomalias identificadas pelas classes abordadas neste trabalho que são: *EAPOLStart, BeaconFlood, Deauthentication, RTSFlood*.

Contudo, técnicas e ferramentas adotadas por gestores de Redes no âmbitos das Redes de Computadores estruturadas nem sempre atendem a estas necessidades de maneira eficiente. Neste sentido a utilização de técnicas de inteligência computacional tornam-se importantes para minimizarem estas dificuldades com o objetivo cada vez mais de identificar anomalias em redes Ad Hoc, através do baixo consumo computacional.

Este trabalho, apresentou um Sistema de Detecção e Classificação em redes Ad Hoc com arquitetura *stand-alone*, de forma a combinar estratégias de Inteligência Computacional para a detecção e classificação de anomalias nestas redes. Estas estratégias são caracterizadas pelos algoritmos inteligentes, que são capazes de minimizar as dificuldades que gestores possuem em controlar os diversos integrantes destas redes, bem como na identificação de diversas anomalias. O Sistema de Detecção e Classificação proposto é uma ferramenta que colabora com a política de segurança em redes Ad Hoc, sendo que para garantir a efetividade da segurança nestes ambientes é importante a associação desta com outras técnicas de segurança.

Os algoritmos inteligentes presentes no Sistema de Detecção e Classificação proposto são técnicas de classificação, as quais possuem a característica de aprendizagem supervisionada e não supervisionada. O sistema proposto é dividido em duas etapas. A primeira destina-se ao agrupamento dos dados de auditoria capturados da rede Ad Hoc, através da utilização do algoritmo de aprendizagem não supervisionada K-Médias. A segunda etapa, então é responsável pela classificação dos dados, através da aprendizagem supervisionada utilizando a Rede Neural Artificial do tipo *MultiLayer Perceptron*. Para a utilização desta Rede Neural há previamente um treinamento de dados da rede Ad Hoc reconhecendo as classes pré-definidas.

O Sistema de Detecção de Intrusão e Classificação proposto tem como saída informações para cada quadro de dados da rede Ad Hoc analisado, sendo que caso seja classificado corretamente então as informações do quadro são inseridas no *log* do sistema. Entretanto, caso o quadro de dados da rede Ad Hoc não seja classificado corretamente é armazenado estas informações sepa-

radamente para posterior análise se for necessário. Desta forma pode-se afirmar que gestores de ambientes de redes Ad Hoc têm a possibilidade de verificar localmente anomalias nas comunicações de cada dispositivo pertencentes às redes Ad Hoc.

O processo de validação dos algoritmos selecionados envolve a utilização de uma base de dados real de tráfego de redes Ad Hoc, a qual possui injeção das seguintes anomalias: *EAPOLStart, BeaconFlood, Deauthentication e RTSFlood*. Esta base de dados é composta por 17 variáveis por cada quadro, sendo 16 atributos da camada MAC e um atributo de identificação da classe a qual pertence determinado quadro. Também utiliza-se a base de dados KDD 99 para validação dos algoritmos, sendo identificado as seguintes classes de anomalias: *U2R, R2L, Probe, DoS*.

Os resultados obtidos permitem recomendar a Rede Neural *MultiLayer Perceptron* com aprendizado *backpropagation* para a classificação de anomalias em ambientes de Redes Sem Fio, pois possui uma taxa de acertos na classificação de 93,35% dos registros para a base de dados *Wireless* e Ad Hoc e uma taxa de classificação de 99,80% para a base de dados KDD 99. São classificadas todas as classes predefinidas nas mesmas(*Normal, EAPOLStart, BeaconFlood, Deauthentication, RTSFlood, U2R, R2L, Probe, DoS*). O algoritmo K-Médias é sugerido pois consegue realizar o agrupamento dos dados em 61,78 segundos para a base de dados *Wireless* e Ad Hoc e 43,20 segundos para a base de dados KDD 99, sendo que após um processo de rotulação de cada *cluster*, pode-se verificar um acerto de 95,35% e 99,60% para as respectivas bases de dados.

Diante dos resultados obtidos com a Rede Neural *MultiLayer Perceptron* e o algoritmo *K-Médias*, estes são definidos para atuarem no Sistema de Detecção e Classificação para redes Ad Hoc proposto neste trabalho. Esta escolha se concretiza após a análise dos resultados obtidos e também pela presença destes algoritmos inteligentes em propostas que envolvam Sistemas de Detecção de Intrusão em ambientes de Redes Sem Fio.

A validação do Sistema de Detecção e Classificação de Intrusão proposto utiliza a base de dados real de redes Ad Hoc, em que primeiramente captura-se os dados de auditoria para organizá-los em grupos ou *clusters* para posterior classificação. O algoritmo K-Médias, no entanto, organiza os dados em 10 *clusters* em 19 iterações em 14,82 segundos, confirmando sua principal característica de baixo poder computacional. Assim, a Rede Neural *MultiLayer Perceptron*, realiza o processo de classificação possuindo taxa de acerto em 98,07%, com erro médio absoluto de 0,0316, média da raiz quadrada do erro de 0,127% e coeficiente *Kappa* de 0,8491. O sistema proposto é competente para o processo de classificação de dados de redes sem fio Ad Hoc, através da utilização de estrutura em etapas, pois apresenta uma taxa de falso positivos em 0,5% e falso negativos em 12,36%, contribuindo para viabilidade de sua taxa de classificação.

A partir dos dados obtidos na validação do Sistema de Detecção e Classificação de Intrusão proposto, juntamente com as fundamentações teóricas de Sistemas de Detecção de Intrusão, bem como, das técnicas de inteligência computacional adotadas, afirma-se que o sistema proposto se torna viável para o processo de classificação e identificação de anomalias em quadros de redes sem fio Ad Hoc. Deve-se salientar que os dispositivos móveis utilizados nestas avaliações tiveram um bom comportamento, sendo que a eficiência de consumo de energia ainda é um problema para

aplicações em redes sem fio Ad Hoc. Assim, o Sistema de Detecção e Classificação proposto não garante a segurança total de ambientes de redes Ad Hoc, sendo necessária a utilização de demais tecnologias em conjunto para prover um ambiente estável.

É esperado que a maior parte do tempo, uma rede Ad Hoc funcione em condição normal, que os ataques aconteçam durante pequenos períodos de tempo. É interessante então haver um mecanismo que indique condições anômalas na rede, mas que demande baixo poder computacional. Por isso, a abordagem proposta com o uso do algoritmo K-Médias e Redes Neurais Artificiais *MultiLayer Perceptron* é a principal contribuição desta tese.

Os trabalhos futuros podem ser realizados de diversas maneiras:

- Aplicação em Redes Sem Fio, avaliando o sistema proposto em uma rede sem fio real em seus diversos ambientes, desde corporativos até em ambientes de desastres. Serão esperados ataques específicos dessa arquitetura, sendo necessário o uso de atributos obtidos na camada de enlace para detecção e classificação dos ataques;
- Estudo da adaptação deste sistema proposto para avaliações instantâneas na rede, ou seja, possibilidade de detecção *on line*;
- Estudo e análise da utilização de outras técnicas de inteligência computacional, com o objetivo de acelerar o processamento do sistema proposto, sem grandes prejuízos para os dispositivos das redes Ad Hoc;
- Integração com Sistema de prevenção de ataques, pois o sistema proposto indica se determinado dispositivo está sob condição normal ou de anomalia. Se houver uma situação anômala, o administrador deverá realizar ações para anular ou minimizar os danos. Assim, poderá ser realizada a integração com um sistema de prevenção de ataques, sendo que ações automáticas poderão ser executadas sem a intervenção manual de um funcionamento correto da rede;
- Realização da análise semântica das informações pertencentes aos quadros classificados de maneira correta, podendo o administrador ter acesso a informações mais específicas do tráfego anômalo da Rede;
- Elaboração de uma nova base de dados constituída de tráfego de redes de computadores contendo protocolos novos, bem como com anomalias atuais;
- Utilização desta proposta para a classificação e predição por exemplo de mensagens textuais, através da possibilidade de análise semântica dos dados capturados e pre-processados presentes no sistema proposto.

Referências Bibliográficas

1 CERT-BR. *Estatísticas dos Incidentes Reportados ao CERT.br*. <https://www.cert.br/stats/incidentes/>. Acesso em: 01 junho 2018.

2 FERREIRA, E. W. T. *Proposta de um sistema de detecção e classificação de intrusão em redes de computadores baseado em transformadas wavelets e redes neurais artificiais*. Dissertação (Tese) — Universidade Federal de Uberlândia, Uberlandia, Brasil, Dezembro 2009.

3 GARCIA, G. *Avaliação de um protocolo MAC sem fio full duplex*. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, 2016.

4 SGORA, A.; VERGADOS, D. D.; CHATZIMISIOS, P. A survey on security and privacy issues in wireless mesh networks. *Security and Communication Networks*, Wiley Online Library, v. 9, n. 13, p. 1877–1889, 2016.

5 FIORIN, D. V.; MARTINS, F.; PEREIRA, E.; SCHUCH, N. J. Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. v. 33, p. 1309, 03 2011.

6 HAYKIN, S. O. *Neural Networks and Learning Machines*. 3nd. ed. Ontario Canada: Pearson, 2009.

7 ACADEMY, D. S. *Deep Learning Book*. <http://deeplearningbook.com.br/>. Acesso em: 01 junho 2018.

8 FERREIRA, E. W. T.; SHINODA, A. A.; OLIVEIRA, R. D.; NASCIMENTO, V. E.; ARAúJO, N. V. D. S. A Methodology for building a Dataset to Assess Intrusion Detection Systems in Wireless Networks. *WSEAS Transactions on Communications*, v. 14, p. 113–120, 2015.

9 AMARAL, R. P. F.; BARBOSA, F. de S.; CURY, A. A.; FONSECA, L. G. da; BONIFÁCIO, A. L. Aplicação de métodos computacionais a dados vibracionais para detecção de alterações estruturais. *XII Simpósio de Mecânica Computacional*, Universidade Federal de Juiz de Fora (UFJF), p. 186–191, 2016.

10 BAID, A.; RAYCHAUDHURI, D. Understanding channel selection dynamics in dense wi-fi networks. *IEEE Communications Magazine*, IEEE, v. 53, n. 1, p. 110–117, 2015.

11 CANEDO, D. R.; ROMARIZ, A. R. S. Data analysis of wireless networks using computational intelligence. *Journal of Communications*, v. 13, n. 11, p. 618–626, 2018.

12 SHA, M.; GUNATILAKA, D.; WU, C.; LU, C. Empirical study and enhancements of industrial wireless sensor–actuator network protocols. *IEEE Internet of Things Journal*, IEEE, v. 4, n. 3, p. 696–704, 2017.

13 LOO, J.; MAURI, J. L.; ORTIZ, J. H. *Mobile ad hoc networks: current status and future trends*. [S.l.]: CRC Press, 2016.

14 PUTTINI, R. S. *Um Modelo de Segurança para Redes Móveis Ad Hoc*. Dissertação (Tese) — Universidade de Brasília, Brasília, Brasil, Outubro 2004.

15 GOGOI, P.; BHATTACHARYYA, D.; BORAH, B.; KALITA, J. K. Mlh-ids: A multi-level hybrid intrusion detection method. *The Computer Journal*, v. 57, n. 4, p. 602–623, 2014. Disponível em: <http://dx.doi.org/10.1093/comjnl/bxt044>.

16 CHANDRASHEKAR, G.; SAHIN, F. A survey on feature selection methods. *Computers & Electrical Engineering*, v. 40, n. 1, p. 16–28, 2014.

17 BHATTACHARYA, S.; SELVAKUMAR, S. Multi-measure multi-weight ranking approach for the identification of the network features for the detection of dos and probe attacks. *The Computer Journal*, v. 59, n. 6, p. 923–943, 2016. Disponível em: <http://dx.doi.org/10.1093/comjnl/bxv078>.

18 VLĂDUŢU, A.; COMĂNECI, D.; DOBRE, C. Internet traffic classification based on flows' statistical properties with machine learning. *International Journal of Network Management*, Wiley Online Library, v. 27, n. 3, p. e1929, 2017.

19 NISHANI, L.; BIBA, M. Machine learning for intrusion detection in manet: a state-of-the-art survey. *Journal of Intelligent Information Systems*, Springer, v. 46, n. 2, p. 391–407, 2016.

20 GOVINDARAJAN, M.; CHANDRASEKARAN, R. Intrusion detection using neural based hybrid classification methods. *Computer Networks*, v. 55, n. 8, p. 1662–1671, 2011.

21 FERREIRA, E. W. T.; CARRIJO, G. A.; OLIVEIRA, R. de; ARAUJO, N. V. de S. Intrusion detection system with wavelet and neural artifical network approach for networks computers. *IEEE Latin America Transactions*, IEEE, v. 9, n. 5, p. 832–837, 2011.

22 VO, V.; LUO, J.; VO, B. Time series trend analysis based on k-means and support vector machine. *Computing and Informatics*, v. 35, p. 11–127, 2016. Disponível em: <https://pdfs.semanticscholar.org/fd6d/6d3778f52608f048aa95dd9aaca42fe2871f.pdf>.

23 NUNES, B. A. A.; MENDONCA, M.; NGUYEN, X.-N.; OBRACZKA, K.; TURLETTI, T. A survey of software-defined networking: Past, present, and future of programmable networks. *IEEE Communications Surveys & Tutorials*, IEEE, v. 16, n. 3, p. 1617–1634, 2014.

24 RUSSELL, A. L.; SCHAFER, V. In the shadow of arpanet and internet: Louis pouzin and the cyclades network in the 1970s. *Technology and Culture*, The Johns Hopkins University Press, v. 55, n. 4, p. 880–907, 2014.

25 CALMON, F. du P.; CLOUD, J. M.; MEDARD, M.; ZENG, W. *Multi-path data transfer using network coding*. [S.l.]: Google Patents, 2017. US Patent 9,537,759.

26 KUROSE, J. F. e. K. W. R. *Redes de Computadores e a Internet*. [S.l.]: Pearson Education Companion, 2006. (Pearson Education Companion).

27 ZHANG, H.; LOU, H.-L.; NABAR, R. U.; SRINIVASA, S.; YU, M.; BANERJEA, R. *Physical layer frame format for WLAN*. [S.l.]: Google Patents, 2017. US Patent 9,655,002.

28 ZHANG, Z.; LONG, K.; WANG, J.; DRESSLER, F. On swarm intelligence inspired self-organized networking: its bionic mechanisms, designing principles and optimization approaches. *IEEE Communications Surveys & Tutorials*, IEEE, v. 16, n. 1, p. 513–537, 2014.

29 MISHRA, V.; MATHEW, J.; LAU, C.-T. Introduction. In: *QoS and Energy Management in Cognitive Radio Network*. [S.l.]: Springer, 2017. p. 1–37.

30 CHEN, J.; YU, Q.; CHAI, B.; SUN, Y.; FAN, Y.; SHEN, X. S. Dynamic channel assignment for wireless sensor networks: A regret matching based approach. *IEEE Transactions on Parallel and Distributed Systems*, IEEE, v. 26, n. 1, p. 95–106, 2015.

31 GAVRILOVSKA, L.; DENKOVSKI, D.; RAKOVIC, V.; ANGJELICINOSKI, M. Medium access control protocols in cognitive radio networks. In: *Cognitive Radio and Networking for Heterogeneous Wireless Networks*. [S.l.]: Springer, 2015. p. 109–149.

32 PETITE, D. *Wireless network protocol systems and methods*. [S.l.]: Google Patents, 2018. US Patent 9,860,820.

33 PERANCONI, D. S.; MUHAMMAD, H. H.; BARCELLOS, M. P. Modelo de arquitetura para simulação de redes móveis sem fio ad hoc no simmcast. *Unisinos-Universidade do Vale do Rio dos Sinos*, p. 4, 2011.

34 MUKHERJEE, A.; FAKOORIAN, S. A. A.; HUANG, J.; SWINDLEHURST, A. L. Principles of physical layer security in multiuser wireless networks: A survey. *IEEE Communications Surveys & Tutorials*, IEEE, v. 16, n. 3, p. 1550–1573, 2014.

35 SARKAR, S. K.; BASAVARAJU, T.; PUTTAMADAPPA, C. *Ad hoc mobile wireless networks: principles, protocols, and applications*. [S.l.]: CRC Press, 2016.

36 ARAÚJO, G.; FERREIRA, G.; PONTES, D. N. Tecnologia móvel e ubíqua na educação: uma análise com base em revisão sistemática. In: *Anais do Workshop de Informática na Escola*. [S.l.: s.n.], 2017. v. 23, n. 1, p. 1052.

37 MENDONçA, I. E. S. *Explorando Funcionalidades Sociais e Colaborativas Em Ambientes Educacionais Ubíquos*. Dissertação (Dissertação) — Universidade Federal de Uberlândia, Programa de Pós-Graduação em Ciência da Computação, Uberlândia, Brasil, Maio 2015.

38 CHAGAS, A. L. R. Rede veicular: Vanet hibrida, 3g e 4g e geoposicionamento a serviço da segurança pessoal. Universidade Federal do Rio de Janeiro, 2016.

39 PATHAN, A.-S. K. *Security of self-organizing networks: MANET, WSN, WMN, VANET*. [S.l.]: CRC press, 2016.

40 JAHANSHAHI, M.; BARMI, A. T. Multicast routing protocols in wireless mesh networks: a survey. *Computing*, Springer, v. 96, n. 11, p. 1029–1057, 2014.

41 MARLON, d. S.; SENNE, E. L. F.; VIJAYKUMAR, N. L. An optimization model to minimize the expected end-to-end transmission time in wireless mesh networks. *Sobrapo*, Scielo, v. 37, n. 2, p. 209 – 227, 2017.

42 FORCE, I. E. T. *Internet standards*. <https://www.ietf.org/standards/>. Acesso em: 23 jan. 2019.

43 MARINS, A. X. d. *Protocolos de roteamento para redes móveis comparativo : OLSR X AODV*. Dissertação (Projeto Final) — Universidade Federal Fluminense, Niterói, Brasil, 2017.

44 GODINHO, R. D. C. *Levantamento de problemas de segurança nas redes sem fios*. Dissertação (Dissertação) — Faculdade de Ciências - Universidade do Porto, Porto, Protugal, 2016.

45 CUTRIM, C. M. d. O. *Segurança em Redes: Segurança em Redes Sem Fio*. Dissertação (Trabalho de Conclusão de Curso (Especialista em Segurança da Informação) — SENAC - DF, Brasília, Brasil, 2013.

46 GHILDIYAL, S.; MISHRA, A. K.; GUPTA, A.; GARG, N. Analysis of denial of service (dos) attacks in wireless sensor networks. *IJRET: International Journal of Research in Engineering and Technology*, v. 3, p. 2319–1163, 2014.

47 KALANTARY, S.; TAGHIPOUR, S. A survey on architectures, protocols, applications, and management in wireless sensor networks. *Journal of Advanced Computer Science & Technology*, Science Publishing Corporation, v. 3, n. 1, p. 1, 2014.

48 WEI, Z.; TANG, H.; YU, F. R.; WANG, M.; MASON, P. Security enhancements for mobile ad hoc networks with trust management using uncertain reasoning. *IEEE Transactions on Vehicular Technology*, IEEE, v. 63, n. 9, p. 4647–4658, 2014.

49 OLIVEIRA, T. A. *Redes Dinâmicas de Sensores Sem Fio ZigBee para Aplicações de Monitoramento e Controle*. Dissertação (Dissertação) — Universidade Estadual Paulista, Bauru, Brasil, Outubro 2015.

50 GOIS, D. A. S.; LIMA, J. P. A.; ORDONEZ, M.; DAVID, E. Segurança em redes de sensores sem fio-desafios, tendências e orientações. *GESTÃO. Org: Revista Eletrônica de Gestão Organizacional*, v. 13, 2015.

51 MORETTI, C.; BELLEZI, M. A. Segurança em redes sem fio 802.11. *Revista TIS*, v. 3, n. 1, 2014.

52 EVANGELISTA, D.; SILVA, E. da; NOGUEIRA, M.; SANTOS, A. Um controle de associações resistente a ataques sybil para a disseminação segura de conteúdo da iot. *XVI Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais*, 2016.

53 MORAES, R. F. D.; ARAÚJO, N. V. de S.; MACIEL, C. Avaliação de um conjunto de dados quanto à sua qualidade na especificação de perfis de ataque e não-ataque numa rede ieee 802.11 w. *Anais da Escola Regional de Informática da Sociedade Brasileira de Computação (SBC)–Regional de Mato Grosso*, v. 6, p. 145–150, 2015.

54 KOLIAS, C.; KOLIAS, V.; KAMBOURAKIS, G. Termid: a distributed swarm intelligence-based approach for wireless intrusion detection. *International Journal of Information Security*, Springer, v. 16, n. 4, p. 401–416, 2017.

55 JABEZ, J.; MUTHUKUMAR, B. Intrusion detection system (ids): anomaly detection using outlier detection approach. *Procedia Computer Science*, Elsevier, v. 48, p. 338–346, 2015.

56 WHITE, G. B.; FISCH, E. A.; POOCH, U. W. *Computer system and network security*. [S.l.]: CRC press, 2017.

57 BUCZAK, A. L.; GUVEN, E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, IEEE, v. 18, n. 2, p. 1153–1176, 2016.

58 SARIKA, S.; PRAVIN, A.; VIJAYAKUMAR, A.; SELVAMANI, K. Security issues in mobile ad hoc networks. *Procedia Computer Science*, Elsevier, v. 92, p. 329–335, 2016.

59 KUMAR, S.; DUTTA, K. Intrusion detection in mobile ad hoc networks: techniques, systems, and future challenges. *Security and Communication Networks*, Wiley Online Library, v. 9, n. 14, p. 2484–2556, 2016.

60 SPANOS, D. *Intrusion Detection Systems for Mobile Ad Hoc Networks*. Dissertação (Dissertação) — School of Science Technology - International Hellenic University, Thessaloniki, Greece, 2017.

61 ROOPA, M.; RAJA, S. S. Artificial neural network using back propagation algorithm in distributed manets. In: IEEE. *2016 International Conference on Information Communication and Embedded Systems (ICICES)*. [S.l.], 2016. p. 1–4.

62 ZHANG, B.; LIU, Z.; JIA, Y.; REN, J.; ZHAO, X. Network intrusion detection method based on pca and bayes algorithm. *Security and Communication Networks*, Hindawi, v. 2018, 2018.

63 SHAMS, E. A.; RIZANER, A. A novel support vector machine based intrusion detection system for mobile ad hoc networks. *Wireless Networks*, Springer, v. 24, n. 5, p. 1821–1829, 2018.

64 RAJMAHANTY, P. H.; GANAPATHY, S. Role of decision trees in intrusion detection systems: A survey. *International Journal of Advances in Computer and Electronics Engineering*, v. 2, n. 4, p. 09–13, 2017.

65 JUSTIN, V.; MARATHE, N.; DONGRE, N. Hybrid ids using svm classifier for detecting dos attack in manet application. In: IEEE. *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. [S.l.], 2017. p. 775–778.

66 SHONA, D.; KUMAR, M. S. Efficient ids for manet using hybrid firefly with a genetic algorithm. In: IEEE. *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. [S.l.], 2018. p. 191–194.

67 BASSO, M.; VIEIRA, J. P.; PARREIRA, F. J.; SILVEIRA, S. R.; SOUZA, A. S. de. Sistema inteligente para apoio ao diagnóstico de diabetes empregando redes neurais. *Encontro Anual de Tecnologia da Informação e Semana Acadêmica de Tecnologia da Informação*, n. 1, p. 56–63, 2014.

68 SILVA, I. N. D.; SPATTI, D. H.; FLAUZINO, R. A. Redes neurais artificiais para engenharia e ciências aplicadas curso prático. *São Paulo: Artliber*, 2010.

69 QUILICI-GONZALEZ, J. A.; ZAMPIROLLI, F. de A. *Sistemas inteligentes e mineração de dados*. [S.l.: s.n.], 2015.

70 ALVES, M. F.; LOTUFO, A. D. P.; LOPES, M. L. M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, v. 1, n. 1, 2013.

71 GURNEY, K. *An introduction to neural networks*. [S.l.]: CRC press, 2014.

72 MAPANGA, I.; KUMAR, V.; MAKONDO, W.; KUSHBOO, T.; KADEBU, P.; CHANDA, W. Design and implementation of an intrusion detection system using mlp-nn for manet. In: IEEE. *2017 IST-Africa Week Conference (IST-Africa)*. [S.l.], 2017. p. 1–12.

73 TULI, H.; KUMAR, S. Packet delay prediction in manet using artificial neural network. In: *Next-Generation Networks*. [S.l.]: Springer, 2018. p. 369–375.

74 MANTERE, M.; SAILIO, M.; NOPONEN, S. A module for anomaly detection in ics networks. In: ACM. *Proceedings of the 3rd international conference on High confidence networked systems*. [S.l.], 2014. p. 49–56.

75 KOHONEN, T. Self-organizing maps, vol. 30 of springer series in information sciences. *ed: Springer Berlin*, 2001.

76 CABRAL, C. C. *Emprego de técnicas de clusterização para gestão tributária do solo urbano: um estudo de caso aplicado à cidade de Fortaleza*. Dissertação (Dissertação) — Universidade Estadual do Ceará, Fortaleza, Brasil, Maio 2010.

77 BANKOVIC, Z.; FRAGA, D.; MOYA, J. M.; VALLEJO, J. C.; MALAGÓN, P.; ARAUJO, Á.; GOYENECHE, J.-M. de; ROMERO, E.; BLESA, J.; VILLANUEVA, D. et al. Improving security in wmns with reputation systems and self-organizing maps. *Journal of Network and Computer Applications*, Elsevier, v. 34, n. 2, p. 455–463, 2011.

78 WANG, W.; WANG, H.; WANG, B.; WANG, Y.; WANG, J. Energy-aware and self-adaptive anomaly detection scheme based on network tomography in mobile ad hoc networks. *Information Sciences*, Elsevier, v. 220, p. 580–602, 2013.

79 SHINY, X. A.; KANNAN, R. J. Energy efficient clustering protocol using self organizing map in manet. *Indian Journal of Science and Technology*, Indian Society for Education and Environment, v. 8, n. 28, p. 1, 2015.

80 GAVHALE, M.; SARAF, P. D. Survey on algorithms for efficient cluster formation and cluster head selection in manet. *Procedia Computer Science*, Elsevier, v. 78, p. 477–482, 2016.

81 MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. [S.l.], 1967. v. 1, n. 14, p. 281–297.

82 FURLAN, C. P. P. *Análise da Rede Social Tocantins Digital, utilizando o Algoritmo k-médias e centralidade de intermediação*. Dissertação (Dissertação) — Pontifícia Universidade Católica de Goiás, Programa de Mestrado em Engenharia de Produção e Sistemas, Goiânia, Brasil, 2014.

83 ZEEBAREE, D. Q.; HARON, H.; ABDULAZEEZ, A. M.; ZEEBAREE, S. R. M. Combination of k-means clustering with genetic algorithm: A review. *International Journal of Applied Engineering Research*, Research India Publications, v. 12, p. 14238–14245, 2017.

84 KUNDU, A.; MISRA, R.; KAR, A.; DEBCHOUDHURY, S.; PAREEK, S.; NAYAK, S.; DEY, R. On demand secure routing protocol using convex-hull & k-mean approach in manet. In: IEEE. *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. [S.l.], 2016. p. 1–5.

85 POPLI, R.; GARG, K.; BATRA, S. Secham: Secure and efficient cluster head selection algorithm for manet. In: IEEE. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.], 2016. p. 1776–1779.

86 RAO, M.; SINGH, N. Energy efficient qos aware hierarchical kf-mac routing protocol in manet. *Wireless Personal Communications*, Springer, v. 101, n. 2, p. 635–648, 2018.

87 MUTHURAJKUMAR, S.; GANAPATHY, S.; VIJAYALAKSHMI, M.; KANNAN, A. An intelligent secured and energy efficient routing algorithm for manets. *Wireless Personal Communications*, Springer, v. 96, n. 2, p. 1753–1769, 2017.

88 MEENA, G.; CHOUDHARY, R. R. A review paper on ids classification using kdd 99 and nsl kdd dataset in weka. In: IEEE. *2017 International Conference on Computer, Communications and Electronics (Comptelix)*. [S.l.], 2017. p. 553–558.

89 RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533, 1986.

90 HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, ACM, v. 11, n. 1, p. 10–18, 2009.

91 GUENNOUN, M.; LBEKKOURI, A.; BENAMRANE, A.; BEN-TAHIR, M.; EL-KHATIB, K. Wireless networks security: Proof of chopchop attack. In: IEEE. *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*. [S.l.], 2008. p. 1–4.

92 COMBS, G. *Wireshark*. <https://www.wireshark.org/>. Acesso em: 01 junho 2018.

93 SAHU, S. K.; SARANGI, S.; JENA, S. K. A detail analysis on intrusion detection datasets. In: IEEE. *Advance Computing Conference (IACC), 2014 IEEE International*. [S.l.], 2014. p. 1348–1353.

94 JI, H.; KIM, D.; SHIN, D.; SHIN, D. A study on comparison of kdd cup 99 and nsl-kdd using artificial neural network. In: *Advances in Computer Science and Ubiquitous Computing*. [S.l.]: Springer, 2017. p. 452–457.

95 KUSHWAHA, P.; BUCKCHASH, H.; RAMAN, B. Anomaly based intrusion detection using filter based feature selection on kdd-cup 99. In: IEEE. *TENCON 2017-2017 IEEE Region 10 Conference*. [S.l.], 2017. p. 839–844.

96 SADIQ, A. S.; ALKAZEMI, B.; MIRJALILI, S.; AHMED, N.; KHAN, S.; ALI, I.; PATHAN, A.-S. K.; GHAFOOR, K. Z. An efficient ids using hybrid magnetic swarm optimization in wanets. *IEEE Access*, IEEE, v. 6, p. 29041–29053, 2018.

97 AUNG, Y. Y.; MIN, M. M. An analysis of random forest algorithm based network intrusion detection system. In: IEEE. *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017 18th IEEE/ACIS International Conference on*. [S.l.], 2017. p. 127–132.

98 SINGH, D. K. A detail analysis of kdd 1999, nsl kdd 1999 and gurekdd dataset of intrusion detection system. *International Journal of Advanced in Management, Technology and Engineering Sciences*, v. 8, 2018.

99 ALI, M. H.; MOHAMMED, B. A. D. A.; ISMAIL, M. A. B.; ZOLKIPLI, M. F. A new intrusion detection system based on fast learning network and particle swarm optimization. *IEEE Access*, IEEE, 2018.

100 MAHMOOD, H. A. Network intrusion detection system (nids) in cloud environment based on hidden naïve bayes multiclass classifier. *Al-Mustansiriyah Journal of Science*, v. 28, n. 2, p. 134–142, 2018.

101 ZUECH, R.; KHOSHGOFTAAR, T. M. A survey on feature selection for intrusion detection. In: *Proceedings of the 21st ISSAT International Conference on Reliability and Quality in Design*. [S.l.: s.n.], 2015. p. 150–155.

102 MAGALHãES, J. M. d. C. *Classificação de Atributos através do Ganho de Informação para efeitos de Reconhecimento de Browsers*. Dissertação (Dissertação) — Faculdade de Engenharia da Universidade do Porto, 2010.

103 CHOUDHURY, S.; BHOWAL, A. Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection. In: IEEE. *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*. [S.l.], 2015. p. 89–95.

104 ALAEI, P.; NOORBEHBAHANI, F. Incremental anomaly-based intrusion detection system using limited labeled data. In: IEEE. *2017 3th International Conference on Web Research (ICWR)*. [S.l.], 2017. p. 178–184.

105 LAKSHMI, G.; PATIL, S. B.; PATIL, P. Hybrid intrusion detection using a zone based aodv routing protocol for manets. *Journal of Computational and Theoretical Nanoscience*, American Scientific Publishers, v. 15, n. 11-12, p. 3266–3274, 2018.

106 CERVANTES, C.; NOGUEIRA, M.; SANTOS, A. Mitigação de ataques no roteamento em iot densa e móvel baseada em agrupamento e confiabilidade dos dispositivos. In: SBC. *Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. [S.l.], 2018.

A. APÊNDICE(ARTIGO PUBLICADO) - DATA ANALYSIS OF WIRELESS NETWORKS USING COMPUTATIONAL INTELLIGENCE

Data Analysis of Wireless Networks Using Computational Intelligence

Daniel R. Canêdo[1,2] and Alexandre R. S. Romariz[1]
[1] Universidade de Brasília-UnB/Departamento de Engenharia Elétrica, Brasília, Brazil
[2] Instituto Federal de Goiás - IFG, Luziânia, Brazil
Email: daniel.canedo@ifg.edu.br; romariz@ene.unb.br

Abstract**—**In the last decade a great technological advance in mobile technologies infrastructure was seen. The increase in the use of wireless local networks and the use of services from satellites is also noticed. The high utilization rate of mobile devices for various purposes makes clear the need to monitor wireless networks to ensure the integrity and confidentiality of the information transmitted. Therefore, it is necessary to quickly and efficiently identify the normal and abnormal traffic of these networks, so that the administrators can take action. This work aims, from a database of wireless networks, to classify this data in some classes of pre-established anomalies according to some defined criteria of the MAC layer, using supervised and unsupervised intelligent algorithms Multilayer Perceptron (MLP), K-Means and Self-Organizing Maps (SOM). For the analysis of the mentioned algorithms, the WEKA Data Mining software (Waikato Environment for Knowledge Analysis) is used. The algorithms have high success rate in the classification of the data, being indicated in the use of Intrusion Detection Systems for Wireless Networks.

Index Terms**—**Wireless networks, multilayer perceptron, K-means, self-organized map, weka

I. Introduction

In the last decade a great technological advance was seen, especially regarding mobile technologies and its infrastructure. The increase in the use of wireless local networks and also the use of services from satellites, both in organizational and residential environments, is identified. This allows information to be created, transmitted and accessed faster and anywhere at any time by simply having access to the mobile network infrastructure. According to Anatel (Telecommunication National Agency), in January/2016 Brazil registered 257.248 million active lines in mobile telephony, with pre-paid accesses corresponding to 71.45% (183.80 million) of total accesses, while postpaid accesses correspond to 28.55% (73.45 million).

The consequence of this scenario is perceived when the use of computational devices used by both individuals and companies is verified. This scenario can be verified through the research conducted by IDC Brasil, which states that in the last quarter of 2014 Brazil had 1,637 million computers, of which 600 thousand are desktops and 1,037 million are notebooks. An unpublished survey by the Brazilian Institute of Geography and Statistics (IBGE) reveals that 57.3% of homes access the internet through cell phones and tablets in the year 2013.

The Wireless Networks environment, as well as the environment of Ad Hoc Wireless Networks or Wireless Sensor Networks, has in its characteristic a dynamicity in relation to the composition of the network members, that is, for these types of networks users often enter and leave the network. This feature makes real-time management of these environments necessary. This scenario becomes, however, quite vulnerable to attempts to approach the anomalies present in the system as a whole. Anomalies such as *EAPOL Start, Beacon Flood, Deauthentication, RTS Flood* [1].

However, the techniques and tools adopted by network managers in the framework of structured computer networks do not always meet these needs in a timely manner. In this sense, the use of computational intelligence techniques becomes a great option to minimize these difficulties aiming to increasingly identify real-time anomalies.

The high rate of use of mobile devices for various purposes makes clear the need to monitor this infrastructure, since it presents the large-scale transmission of information, which at certain moments may be confidential. The set of this mobile system, determined by both the software and the hardware used, is relatively fragile regarding security, mainly due to the characteristic of its transmission mean, but also due to dynamic access it. So, there is a need to try to quickly and effectively identify the normal and abnormal traffic of these wireless networks so that administrators can take action. This work aims, from a database of wireless networks [1], to classify this data according to some defined criteria of the MAC layer.

The structure of this article is organized into sections. In section two will be presented some works that have the characteristic of identification of wireless networks traffic using algorithms of learning. In section 3, the theoretical basis for Wireless Networks is presented, while section 4 deals with Computational Intelligence Techniques: Neural Networks and K-Means algorithm. Section 5 will present the methodology of experimentation and results. In Section 6 we present the case studies used to analyze the results. In section 7 will be performed the quantitative and qualitative analysis of the results. Section 8 presents the conclusion of the work and future work.

Manuscript received March 22, 2018; revised October 16, 2018.
oi:10.12720/jcm.13.11.618-626

II. Related Works

It is possible to find in the literature some works of Wireless Networks traffic classification, which can be applied in Intrusion Detection Systems. These proposals make use of supervised and unsupervised learning methods. The proposal [2] provides a general approach to the various classification methods, using high-dimensional data and a variable selection technique aiming to reduce computational time and improving the learning rate.

Govindarajan presents a proposal [3] of two classification methods involving multilayer perceptron and Basis function Networks. This work proposes a hybrid architecture involving both classifiers for intrusion detection systems. Ed Wilson presents a proposal [4] of Hybrid Intrusion Detection System, in which signal processing is performed using the Wavelet transform and then the classification of the anomalies using Artificial Neural Networks.

Ed Wilson [1] proposes the elaboration of a real database of Wireless Network traffic, which will be used in the evaluation of Intrusion Detection Systems (IDS). This data, in turn, undergoes a pre-processing to later be classified by techniques of standards recognition, such as Artificial Neural Networks.

III. Wireless Networks

The IEEE 802.11 standard defines an architecture for the Wireless Local Area Network that covers the physical and link levels present in the reference OSI communication model. For the physical level only, Radio Frequency (RF) and infrared (IR) transmissions are treated, but other forms of wireless communication such as microwave and visible light can also be considered. For the link level, the access control to the medium is addressed, through the definition of the MAC protocol (Medium access Control).

Taking into account the main characteristics of the IEEE 802.11 standard, such as interoperability, low cost, high market demand, reliability of project execution, there is a great growth in the use mainly of Local Area Networks of Wireless Computers, also known as Wireless Networks, in public and private environments. This makes Wireless Networks a priority resource in environments where it is most often possible to access the Internet, whether inside corporations, in homes or in public environments, such as shopping malls, airports and so on.

The architecture of Wireless Networks according to the IEEE 802.11 standard, is based on the division of the area covered by the Wireless Network into cells, these cells being called BSA (Basic Service Area). The size of the coverage of each BSA will depend exclusively on the characteristics of the environment itself and the power of transmitters and receivers used in the computational devices. The other components of the Wireless Networks architecture are listed below:

1) BSS (*Basic Service Set*): Which is the set of computational devices that communicate by broadcasting (BC) or infrared (IR) within a Basic Service Area;
2) AP (*Access Point*): Specific computational devices, which have the purpose of capturing the transmissions made by computational devices belonging to its BSA (Basic Service Area), which are destined to stations belonging to another Basic Service Area. The Access Point, in turn, will perform the retransmission using a distribution system;
3) Distribution System: Communication infrastructure, which has the purpose of performing the interconnection of several Basic Service Area to allow the construction of networks, which have covers larger than one cell;
4) ESA (*Extended Service Area*): Service Area that has the purpose of interconnecting several BSAs, through the Distribution System using the Access Point;
5) ESS (*Extended Service Set*): Which is intended to represent a set of computational devices consisting of the union of several BSSs (Basic Service Set) connected by a Distribution System.

The IEEE 802.11 standard also defines a medium access protocol, which is present in a MAC sublayer of the data link level. This protocol is called DFWMAC (*Distributed Foundation Wireless Medium Access Control*), which has two access methods, one of which is a distributed and mandatory feature. The other access method of the DFWMAC protocol is optional, having a centralized feature, and according to the IEEE standard, both the distributed method and the centralized method in the communication system can coexist. The medium access protocol also has the property of treating problems related to computational devices that try to move from one cell to another, a process called roaming. It is also related to the protocol of access to the medium of property to treat problems of lost computational devices, being able to be denominated of hidden node.

IV. Computational Intelligence Techniques

Computational Intelligence consists of an area of computing and engineering responsible for studying the computational principles that make intelligent behavior possible. Among the main techniques of this area are [5]:

- Artificial neural networks;
- Fuzzy Logic;
- Evolutionary Algorithms;
- Theory of Games

A. Artificial Neural Networks

The work related to Artificial Neural Networks, is inspired by the observation that the human brain has unusual computational properties. According to [6], the human brain represents a highly complex, non-linear and parallel information processing system.

An Artificial Neural Network is composed of relatively simple processing units, and the interconnections of these units are adapted from a learning algorithm.

Fig. 1 presents the neuron model present in the Neural Networks, which is composed of input signals, weights, sum function, transfer function and output.

These components are shown below:

- Inputs: It is the signal x_j present at the input of neuron k which is multiplied by the synaptic weight w_{kj}. In the synaptic weight the first index k refers to the neuron in question and the index j refers to the input terminal of the synapse to which the weight is referring.
- Sum function: It is intended to sum the input signals, taking into account the respective synapses of the neuron. The sum function can be obtained by Equation 1:

$$u_k = \Sigma\, w_{kj}\, x_j \tag{1}$$

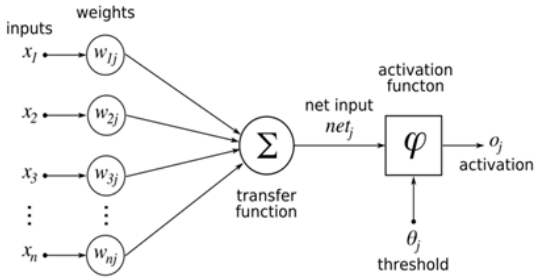


Fig. 1. Model neuron

- Activation Function: Also called "Transfer Function" that maps the sum to an end output.

The sigmoid activation function is the most used function in applications of Artificial Neural Networks, being composed by a S-shaped graph. The sigmoid function is represented in applications of Neural Networks by Equation 2:

$$\varphi(v) = \frac{1}{1+\exp(-av)} \tag{2}$$

In terms of its architecture, Artificial Neural Networks can be classified into two distinct groups: those that do not have recurrent connections, called acyclic ones, and the cyclical ones, which have it [7]. Direct feeding neural networks is organized in layers, and that certain layer can receive only inputs of neurons located in the layer immediately inferior or below. The inner layers (which do not connect to the outside world) are called hidden layers [6].

Acyclic Neural Networks of several layers have the fundamental characteristic of direct feeding, but with the presence of one or several layers hidden between the input layer and the output layer. The differentiation between Neural Networks that make use of hidden layers to those that do not use, for example the Perceptron Network, is the possibility of increasing the capacity of representation of transformations between inputs and outputs of the Neural Network.

B. Neural Networks Learning

The learning process consists in the adaptation of Synaptic values. For learning in Neural Networks there are several algorithms able to perform the adaptation of the parameters, so that after a finite number of iterations can converge to a viable solution.

For learning in Neural Networks there are several algorithms able to perform the adaptation of the parameters so that after a finite number of iterations can converge to a viable solution. The learning algorithm aims to reduce a cost function, usually associated with the error in the system output [6].

According to Rezende [8], the algorithms or learning techniques applied to Artificial Neural Networks can be classified according to three different principles:

- Supervised learning
- Non-supervised learning
- Reinforcement learning

Supervised learning has the purpose of enabling the learning of a given Artificial Neural Network through a set of input and output examples. Since the desired output is known for each example, it is possible to calculate the error and adjust the weights, in order to approximate the answer of the desired answer.

Unsupervised learning also has the purpose of enabling the learning of a particular Neural Network through the processing of a set of information, but without presence of a specialist (teacher), that is, for this type of learning it has not the knowledge of the desired outputs for the inputs entered during the training. The adjustments of the synaptic weights belonging to each entry are performed based on the input values [8].

Fig. 2 demonstrates the behavior of a neuron in the supervised learning process, having as fundamental elements the input vector, hidden neuron layer, output neuron, summation function. The neuron has an output $Y_k(n)$, which is the result of the activation function of the neuron. In Fig. 2 there is the presence of a desired output $d_k(n)$, which is possibly different from the output $Y_k(n)$ of the neuron. In this way the subtraction between the output generated by the neuron $Y_k(n)$ and the desired output $d_k(n)$ results in an error signal $e_k(n)$, which will be returned to the neuron, allowing the adjustment of its respective weights of the input layer, generating new outputs, as defined in the equations below:

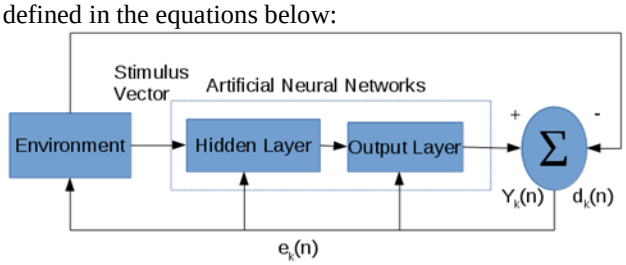


Fig. 2. Learning by error connection.

- The local gradient of the last layer is calculated through the error generated in the output layer and the derivative of the error using equation 3 [9].

$$\delta_j(n) = e_j(n)\varphi'_j(v_j(n)) \quad (3)$$

- Error Propagation: The local gradient of each neuron from the previous layers is calculated by equation 4 [9].

$$\delta_j(n) = \varphi'_j(v_j(n))\Sigma_k\delta_k(n)w_{kj}(n) \quad (4)$$

where:
j: index of the current layer neuron;
k: neuron index of the posterior layer.

- Synaptic weights: After calculating each local gradient, the adjustment of the synaptic weights is given by equation 5 [9].

$$\Delta w_{ji}(n) = \eta\varphi_j(n)y_j(n) \quad (5)$$

where:
η: network.

- The adaptation of the weights is performed by equation 6 [9].

$$w^{k+1}{}_{ji}(n) + \Delta w_{ji}(n) \quad (6)$$

- For each pattern presented to the network, the instantaneous error in equation 7 [9] is measured.

$$\varepsilon(n) = 1/2\Sigma_{j \in C}\, ej^2(n) \quad (7)$$

where:
C: set of all neurons in the output layer

C. Multilayer Perceptron Neural Networks

The Multilayer Perceptron is an Artificial Neural Network architecture composed of an input layer, an output layer and at least one hidden layer between input and output. This type of Neural Network is used in a large scale to solve complex problems, as it has as a characteristic the supervised learning, through the use of the backpropagation error or backpropagation algorithm [6].

The backpropagation algorithm, also known as backpropagation, is used in artificial neural networks with supervised architecture to adjust the synaptic weights, minimizing the errors between the output generated by the neurons and the desired output, that is, to reduce the error rate in each learning cycle [10].

D. Self-Organizing Maps

Artificial Neural Networks called Self-Organizing Maps (*SOM*) are systems that are organized internally to represent the distribution of input data, without the presence of a supervisor. The Kohonem Networks are inspired by the fact that information in the human brain is spatially organized. Areas of the cortex may be said to form "maps" of sensory spaces relating to neurons responsible for specific responses to certain stimuli in certain regions, such as specific responses to frequencies in the areas of the brain intended for hearing and vision. In this sense the Kohonen Networks aim to generate agglomerates with high response activity to a given stimulus.

Self-Organizing Maps (SOM) consist of two layers: The input layer I and the output layer U. The input of the Kohonen Network is a vector in the d-dimensional space in R^d, defined by $x_k = [\xi_1,...,\xi_d]^T$, K = 1, ..., n; each neuron j has a vector w, also represented in the space R^d associated with the input vector x_k, $w_j = [w_{j1}, ..., w_{jd}]^T$. Neurons are interconnected through a neighborhood relation shown in the map structure.

From this, taking into account the state of activation of neuron *i* of the Kohonen Map in relation to the stimulus of the input vector x_k:

1. $y(x_k) = 1$, if $i(x_k) = \arg_j \min\|x_k - w_j\|$;
2. $y(x_k) = 0$, otherwise.

Where:

- $i(x_k)$: i, j represent specific neurons in the network;
- $\|.\|$: It is the distance measure, through the Euclidean norm, $1 \geq j \geq N$. Being N the number of neurons in the output layer;
- $y(x_k)$: Informs the state of activation of the position *i* of the Kohonen Map in relation to the stimulus of the input vector x_k.

The neurons receive the same value the neuron whose weight vector is closer to the input vector get activated. For the winning neuron will be assigned a certain neighborhood. Neurons in this neighborhood will have the opportunity learning, by adapting their weights following Equation 8.

$$w[t + 1] = w[t] + \alpha[t](x[t] - w[t]) \quad (8)$$

where:

- w [t + 1]: Is the value of the weight updated;
- α [t]: It is the learning constant.

Another characteristic in the process of neighborhood determination is to ensure a similarity between the neurons that are part of a given region, this region being composed of the winning neuron and its neighbors. According to Kohonen [11], in order to guarantee similarity, one must apply the adjustments of the weights, as presented in Equation 8, both in the winning neuron, both in its neighborhood. In this way it allows the map to be organized geographically, because the neurons that do not belong to the neighborhood will not have weights adjusted.

The neighborhood concept ensures that close neurons respond to similar patterns, thus creating a self-organizing feature map.

E. K-Means Algorithm

The K-Means algorithm, is a simple technique that can be used to analyze groups. This algorithm is proposed by Macqueen in 1967. The algorithm is applied when grouping certain objects into groups called clusters. According to Macqueen [12] these groups or clusters are

formed through the application of distance measurement techniques or similarity techniques between objects.

K-Means uses a partitioning technique, in which the grouping is performed through optimization through the application of an objective function. This objective function is based on prototypes that have the principle of finding *n* clusters *k*, being the value of n determined by the user [13].

The non-hierarchical grouping process therefore defines a number *k* of classes and also performs an initial classification of n objects in *k* classes, being the value of *k* determined by the user before or after the grouping process.

In terms of programming and computational processing, the K-Means algorithm is easy to program and economical, not requiring high computational power. The K-Means algorithm is able to process large volumes of data, the storage complexity of which is O((m + K) n), where m is the number of points and n is the number of attributes [12].

However, the K-Means algorithm has some disadvantages that need to be analyzed beforehand which are:

- In a large database, the K-Mean algorithm cannot be efficient in generating quality solutions if its initialization is not successful, as well as its initial centroids representing the groups being poorly positioned in the search spacing;
- In terms of performance, the algorithm does not guarantee the optimum overall result, since the final quality of the solution depends on the initial sets of clusters, and can remove them from the overall optimum result;
- Inappropriate choice of the value of k can result in poor results.

The K-Means algorithm therefore has the purpose of converging to a solution using combinations of proximity functions as well as types of centroids reaching a state in which no point, or data object, changes group, consequently there will be no change of centroid. In some cases, the algorithm may not achieve these aims, and it is necessary to assign a weaker condition to reach the final state, such as repeating this process until only 1% of the objects do not change groups.

V. Methodology

This work aims to apply Computational Intelligence techniques to the problem of identifying anomalies in wireless network traffic, more specifically, neural networks and k-means algorithm. As mentioned in the previous sections, the techniques adopted for this work are: Artificial Neural Networks, more precisely with the Multilayer Perceptron algorithm and the Self-Organizing Map (SOM) algorithm, and the K-Means algorithm.

In order to achieve the proposed aims, the following activities were performed in accordance with the chronological order of execution.

We use a database with examples of specific anomalies in wireless networks.. This base in turn is the final product of the work entitled *A Methodology for building a Dataset to Assess Intrusion Detection Systems in Wireless Networks* [1].

The next step is to perform a pre-processing in the database, in such a way that two new databases are obtained. One of the databases is composed of only 10% of the data from the original database and is destined for the test step in the selected algorithms. The other database is composed of 90% of the data from the original database and is destined for the training step of the selected algorithms. Both databases are stored in the Database Manager System named PostgreSQL, and are accessed by the Weka software (*Waikato Environment for Knowledge Analysis*).

Finally, the results of each selected algorithm are analyzed and formatted through tables. In relation to the results, the following information is presented for analysis: Percentage of Classification, labeling of groups, relation of correctness and errors.

VI. Case Study

The case study chosen to analyze the results of the application of computational intelligence techniques presented in previous sections uses data from real wireless networks [1] and the data mining software, Weka [14].

A. Database

The database defined for the execution of this case study is a real collection of network traffic captured in the Wireless architecture. This data, in turn, is obtained by the behavior of users to access various information as well as for the use of the Internet. According to the authors [1], the network traffic obtained by students and employees of the institution in which the experiment was performed was used for this database.

The database chosen for the experimentation of this work made use of two different scenarios. The scenarios discussed have their own configuration and topologies, being a scenario of home environment typical of wireless networks, while the other is a more complex environment, being a corporate environment.

This database is composed of a total of 616,047 records, each record being composed of 16 variables that are characteristics of the wireless network traffic itself. Also in each record of the database is defined a last variable the class to which belongs certain registry, classification is realized taking into account the values of the sixteen variables referring to the obtained wireless network traffic. In this way the data are classified in:

- *Normal*: Acceptable wireless network traffic;
- *EAPOLStart*: Traffic using the Extensible Authentication Protocol (EAP), which aims to perform an authentication method in both the Wired Equivalent Providence (WEP) protocol, both Wi-Fi

Protected Access (WPA) protocol, commercial versions for wireless network access;

- *Beacon Flood*: Management type requests, which are intended to transmit millions of invalid Beacons, resulting in the difficulty that a certain Wireless network device will have in identifying a legitimate Access Point [15];
- *Deauthentication*: It also represents management-type requests, which are injected from the Wireless Network. The frames belonging to this anomaly are transmitted as fictitious requests, which requests the deactivation of a device that is authorized in the Wireless Network;
- *RTSFlood*: Also called Request-to-Send Flood is a control-type frame. This anomaly is based on the large-scale transmission of RTS frames or frames for a short period of time [15].

The database for the experimentation process of this work is divided into two distinct bases, in order to meet the requirements of each defined intelligent algorithm. In this way a training database is generated respecting the characteristics of each algorithm, being composed by 554,442 registers, which corresponds to 90% of the complete database. Also, the test database is generated, being composed by 61,604 records that correspond to 10% of the complete database, respecting the characteristic of each algorithm. In order to optimize the experimentation process and to provide better data manipulation, the training and test databases for each defined computational intelligence technique are stored in the PostgreSQL Database Management System.

B. Experiment 1 – Multilayer Perceptron Algorithm

The Multilayer Perceptron algorithm is one of the algorithms of the Artificial Neural Network that has the characteristic of being supervised, that is, it requires the presence of a specialist in the learning process.

For the realization of the experiment, the Weka software [14] is used through the classification process. In this classification step, one must select the classification algorithm called Multilayer Perceptron, which is assigned properties for its execution:

- Hidden layers of the network are used;
- Learning rate does not decrease with the growth of number of times;
- The number of times to train through.

In order to perform the tests of the Multilayer Perceptron algorithm, the training and testing databases are stored in the PostgreSQL software database. The validation of the algorithm is performed through the use of Cross-Validation, Percentage Separation and Testing techniques.

The cross-validation technique has the characteristic of dividing the database into 10 subsets, in which 9 sets are used for training and one for evaluation. In order to perform this evaluation of the Multilayer Perceptron algorithm we use the complete database, which is composed by 616,047 records.

Percentage separation or Percentage split is a test procedure that has the characteristic of using 66% of the training base, the rest being used for the tests. Also in this case, the complete database (616,047 records) is used for the validation of the Multilayer Perceptron algorithm.

The supplied test set or Supplied test set is a test procedure that makes use of two distinct databases, one for supervised learning of the artificial neural network, while the other database is intended for testing.

Table I presents the percentage of detection of the classes for each of the test procedures performed, while Table II presents the percentage of Registers classified correctly and incorrectly by the Multilayer Perceptron algorithm for each executed test procedure.

TABLE I: ACCURACY – MULTI-LAYER PERCEPTRON

	Cross-Validation	Percentage Split	Supplied Test
Normal	84,97%	87,31%	96,25%
EAPOLStart	4,17%	4,30%	0,81%
BeaconFlood	1,58%	0,93%	0,92%
Deauthentication	2,61%	2,54%	26,65%
RTSFlood	0,02%	0,02%	40%

TABLE II: PERCENTAGE OF ERRORS – MULTI-LAYER PERCEPTRON

	Accuracy	Errors
Cross-Validation	93,34%	6,65%
Percentage Split	95,10%	4,89%
Supplied Test	95,53%	4,46%

C. Experiment 2 – Self-Organizing Maps

Self-Organizing Maps, also called Kohonen Maps is a type of Artificial Neural Networks that have as fundamental principle the competitive procedure of learning between the units of the network.

The main purpose of Kohonen Maps is the possibility of building systems that are organized internally through the distribution of incoming data without the presence of a particular expert. In this sense, Self-Organizing Maps will present in the output the formation of settlements, also called clusters, which have a maximum response to a given stimulus.

In this classification step, it must select the classification algorithm called SOM, which is assigned properties for its execution:

- Initialization of the input vector;
- Define the learning function during training;
- Define the neighborhood function;
- Initialize the initial size of the neighborhood;
- Define the number of training interactions.

In order to perform the tests of the SOM algorithm, the same test options assigned to the Multilayer Perceptron algorithm experiment will be used, which are: Cross-validation Folds 10, Percentage split 66% and Supplied Test Set. In order to perform the Cross-Validation and Percentage Split tests, the complete database (616,047

records) is used, whereas for the Supplied Test Set option, a training database (90% of the complete database) and a base (10% of the complete database).

Table III presents the percentage of detection of the classes for each of the test procedures performed, while Table IV presents the percentage of Records classified correctly and incorrectly by the SOM algorithm for each test procedure performed.

TABLE III: PERCENTAGE OF CLASSIFICATION – SOM

	Cross-Validation	Percentage Split	Supplied Test
Normal	88,07%	88%	88,04%
EAPOLStart	0,59%	0,54%	1,04%
BeaconFlood	0%	0%	0%
Deauthentication	0%	0%	0%
RTSFlood	0%	0%	0%

TABLE IV: PERCENTAGE OF ERRORS – SOM

	Accuracy	Error
Cross-Validation	88,65%	11,34%
Percentage Split	88,54%	11,45%
Supplied Test	89,08%	10,91%

D. Experiment 3 – K-Means Algorithm

The K-Means algorithm is a technique that uses K-Mean data clustering, also called *K-means clustering*. This algorithm seeks to find the best division of data into *K* groups C_i, where i = 1, 2, 3, ..., *K*. Thus, we obtain that the total distance between the data of a given group and its respective center is minimized.

The K-Means algorithm follows the steps below:

- At this step, each point represented by a given P is shifted to its respective group, which corresponds to the nearest mean vector;
- The algorithm calculates again the means of the vectors and also performs the distribution of the data in each group;
- This process of reallocating data to new groups, on which the mean vectors are the closest, is performed until all data are in its groups.

In this step of clustering, it must use the algorithm called Simple K-Means, which is assigned properties for its execution:

- Distance calculation function used, such as Euclidean Distance;
- Define the number of iterations of the algorithm;
- Define the ideal number of clusters to be used to achieve a good result.

To perform the K-Means algorithm tests, the following test options will be used: Supplied Test Set. For the Supplied Test Set, a training database (90% of the complete database) and a test database (10% of the complete database) are used. However, because the K-Means algorithm is unsupervised, data pertaining to the class label is not used for training.

A total of 500 iterations and 25 clusters or groups are defined for the K-Means algorithm. Also, the accuracy for each predefined class (Normal, EAPOLStart, Beacon Flood, Deauthentication, RTS-Flood) is calculated, and these data are presented in Table VI.

TABLE V: CLUSTERING – K-MEANS

	Normal	EAPOL Start	Beacon Flood	Deauthentication	RTSFlood
Clusters	22	2	0	1	0

TABLE VI: PERCENTAGE OF ACCURACY AND ERRORS – K-MEANS

	Accuracy	Errors
Normal	98,39%	1,61%
EAPOLStart	78,48%	21,52%
Beacon Flood	0%	100%
Deauthentication	91,76%	8,24%
RTSFlood	0%	100%

Results show that, with the chosen parameters, the k-means algorithm could not identify all predetermined classes.

For applications of Wireless Networks, which requires an identification of anomalies preferably in real time, as well as the frequent practicality of these anomalies, this algorithm partially meets the desired objective for this context [16]. In order to improve the results of the K-Means algorithm, a technique called Variable Selection, also known as Select Feature, is applied.

The technique of selecting variables has as main objective to select the attributes or variables that can effectively contribute to the achievement of a certain algorithm, that is, it will eliminate attributes considered redundant or irrelevant [2].

In order to improve the performance, as well as the own results obtained by the K-Means algorithm, the CfsSubsetEval algorithm for the identification of the evaluator attribute is applied and for the search method the BestFirst algorithm is used. The first algorithm aims to evaluate a subset of variables considering the individual predictability of each resource, as well as the degree of redundancy among them [17]. The second algorithm aims to searching in a space of attributes subsets that are closer to the objective, which will lead to reach the optimal state more quickly [18].

After applying the technique of selection of variables to the database, which contains 17 attributes, are selected two attributes, which are identified by the third and fifteenth attributes, that is, making use of these two attributes it is possible to achieve the objective more quickly and efficiently.

To perform the test using only the two attributes selected, a total of 500 iterations and 25 clusters or groups were defined for the K-Means algorithm. Fourteen *clusters* identify the Normal class, 3 *clusters* identify the EAPOL Start class, 6 *clusters* identify the Beacon Flood class, 1 *cluster* identifies the Deauthentication class, and

no *cluster* identifies the RTSFlood class. The percentage of correctness and errors for each pre-defined class (Normal, EAPOLStart, Beacon Flood, Deauthentication, RTSFlood) is also performed. These data are presented in Tables VII.

TABLE VII: PERCENTAGE OF ACCURACY AND ERRORS – K-MEANS

	Accuracy	Errors
Normal	97,92%	2,08%
EAPOLStart	59,99%	40,01%
Beacon Flood	68,31%	31,69%
Deauthentication	91,95%	8,05%
RTSFlood	0%	100%

VI. DISCUSSION OF THE RESULTS

The results obtained in the application of the tests in the three experiments described in section 6 point to good results in the application of classification techniques, also known as pattern recognition, for data from wireless networks. Table II, Table IV and Table VI represent the percentage of correctness and errors in relation to the database for the *Multilayer Perceptron* classification algorithm, the *Self-Organizing Maps* algorithm with the Learning Vector Quantization (LVQ) method and the unsupervised learning algorithm K-Means respectively.

The algorithm that best meets the identification needs of the classes determined in the data base of this work is the Perceptron Multilayer classification algorithm with the use of backpropagation training. This algorithm is validated by experiment 1, with 95.53% of the records present in the database being correctly identified for the test set. The other two types of tests adopted (Cross-Validation and Percentage Split) also present satisfactory classification results, with 93.34% and 95.10% respectively.

The *Self-Organizing Maps* and *K-Means* algorithms partially attend the problem of identification of the classes adopted in the database. The first one, according to Table III and Table IV, presents correct identification in 89.08% of the registers, with only the *Normal and EAPOLStart* classes being identified, making it possible to use intrusion detection in wireless networks together with other techniques. The second algorithm presents correct identification in 94.69% of the registers, identifying the *Normal, EAPOLStart, Beacon Flood, Deauthentication* classes, and being unable to identify only the class *RTSFlood*. Although the Multilayer Perceptron algorithm is recommended as the algorithm that best meets the identification needs of the determined classes, the K-Means algorithm, after applying the variable selection technique, achieves satisfactory results, considering the organization of the data and its quantization.

VII. CONCLUSION AND FUTURE WORKS

The aim of this work was to evaluate some computational intelligence algorithms capable of identifying or classifying some anomalies found in the wireless networks traffic. These algorithms are able to minimize the difficulties that managers have in controlling the various members of these networks, as well as in real-time identification of various anomalies.

The algorithms present in this work represent classification techniques, which have the characteristic of supervised and unsupervised learning. For the evaluation of supervised classification, the Multilayer Perceptron algorithm with backpropagation learning is used. For the evaluation of the unsupervised classification, also called the clustering process in which no specialist is present, the K-Means algorithm and the algorithm of the Self-Organizing Maps with Learning Vector Quantization (LVQ) method are used.

The validation process of the selected algorithms involves the use of a real database of Wireless Network traffic, which has the following anomalies: EAPOLStart, Beacon Flood, Deauthentication and RTSFlood. This database is composed of 17 variables per record, 16 attributes of the MAC layer and an attribute of identification of the class to which a particular record belongs.

The best result for the classification of anomalies in wireless environments is the Multilayer Perceptron classification algorithm, since it presents a general correct classification rate of 95.53% and 93.34% for the cross-validation test of the collected database and also classifies all predefined classes (Normal, EAPOLStart, Beacon Flood, Deauthentication, RTSFlood). The other algorithms can also obtain a high percentage of correct answers in the classification, but they cannot identify all the predefined classes. Although the *K-Means* algorithm is partially attend the problem of identification of the classes adopted in the database, it is necessary to apply a pre-processing of variable selection, which can be useful when there is no supervised data.

Future work with the intention of continuing the proposal in this article involves the evaluation of other algorithms of pattern recognition with supervised and unsupervised learning techniques. Also, as future work we can verify the recommendation of supervised or unsupervised classification algorithms with wireless network traffic using WPA (*Wi-Fi Protected Access*) and WEP (*Wired Equivalent Privacy*).

REFERENCES

[1] E. W. T. Ferreira, *et al*., "A methodology for building a dataset to assess intrusion detection systems in wireless networks," *WSEAS Transactions on Communications*, vol. 14, pp. 113–120, 2015.

[2] G. C. F. Sahin, "A survey on feature selection methods," *Computers Electrical Engineering*, 2014, pp. 16–28.
[3] M. Govindarajan and R. M. Chandrasekaran, "Intrusion detection using neural based hybrid classification methods," *Computer Networks* vol. 55, no. 8, pp. 1662–1671, 2011.
[4] E. W. T. Ferreira, *et al*., "Intrusion detection system with wavelet and neural artifical network approach for networks computers," *IEEE Latin America Transactions*, vol. 9, no. 5, pp. 832–837, 2011.
[5] S. Shamshirband, *et al*., "An appraisal and esign of a ulti-agent system based cooperative wireless intrusion detection computational intelligence technique," *Engineering Applications of Artificial Intelligence*, vol. 26, pp. 2015–2127, 2013.
[6] S. Haykin, *Neural Networks and Learning Machines*, 3nd. Ontario Canada: Pearson, 2009.
[7] P. N. E. S. J. Russel, *Intelig ência Artificial*, 2nd. Rio de Janeiro: Elsevier, 2004.
[8] S. O. Rezende, *Sistemas Inteligentes Fundamentos e Aplica ções*, 1nd. Barueri - SP - Brasil: Editora Manole, 2003.
[9] M. F. Alves and A. D. P. L. M. L. M. Lopes, "Selecão de vari áveis stepwise aplicadas em redes neurais artificiais para previs ão de demanda de cargas el étricas," *Simp ósio Brasileiro de Automac ão Inteligente*, 2013.
[10] M. Ibnkahla, "Applications of neural networks to digital communications: A survey," *Signal Processing - Special Issue on Emerging Techniques for Communication Terminals*, vol. 80, no. 7, pp. 1185–1215, 2000.
[11] T. Kohonen, *Self-Organizing Maps*, 3nd. Springer-Verlag Berlin Heidelberg, 2001.
[12] J. B. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.
[13] C. P. P. Furlan, "Análise da Rede Social Tocantins Digital, utilizando o Algoritmo k-m édias e centralidade de intermedia ção," Diss. de mestrado. Pontif ícia Universidade Cat ólica de Goi ás, 2014.
[14] M. Hall, *et al*., "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
[15] R. F. de Moraes and N. V. D. S. A. C. Maciel, "Avaliação de um conjunto de dados quanto á sua qualidade na especifica ção de perfis de ataque e n ão-ataque numa rede IEEE 802.11w," *Anais da VI Escola Regional de Inform ática da Sociedade Brasileira da Computa ção(SBC) - Regional de Mato Grosso*, 2015, pp. 145–1508.
[16] A. Dorri and S. R. K. E. Kheirkhah, "Security challenges in mobile ad hoc networks: A survey," *IJCSES*, vol. 6, no. 1, 2015.
[17] M. A. Hall, "Correlation-based feature subset selection for machine learning," Tese de doutorado. Hamilton, New Zealand: University of Waikato, 1998.
[18] R. E. K. D. M. Chickering, "Best-first minimax search," *Atificial Intelligence*, vol. 84, pp. 299–337, 1996.

Daniel R. Canêdo has a degree in Computer Engineering from Pontif ícia Universidade Cat ólica de Goi ás (2003) and a Master's degree in Electrical Engineering from the University of Bras ília (2006). He is currently an exclusive professor of the Federal Institute of Goi ás - Campus Luzi ânia. He is currently a PhD student in the Post-Graduate Program in Electronic Systems and Automation Engineering of the Department of Electrical Engineering of the University of Bras ília (UnB).

Alexandre R. Romariz holds a BS in Electrical Engineering from the University of Bras ília (1992), a Master's degree in Electrical Engineering from the State University of Campinas (1995) and a PhD in Electrical Engineering from the University of Colorado at Boulder (2003). He is currently "Professor Associado" at the University of Brasilia. He has experience in Computational Intelligence, Integrated Circuits, Optoelectronics and Digital Signal Processing.

B. APÊNDICE(ARTIGO ACEITO - 9TH INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE, ENGINEERING AND APPLICATIONS (CCSEA 2019)) - DATA ANALYSIS OF WIRELESS NETWORKS USING CLASSIFICATION TECHNIQUES

DATA ANALYSIS OF WIRELESS NETWORKS USING CLASSIFICATION TECHNIQUES

Daniel Rosa Canêdo[1,2] and Alexandre Ricardo Soares Romariz[1]

[1]Department of Electrical Engineering, University of Brasília, Brasília, Brazil
daniel.canedo@ifg.edu.br , alromariz@gmail.com
[2]Federal Institute of Goiás, Luziânia, Brazil
daniel.canedo@ifg.edu.br

ABSTRACT

In the last decade, there has been a great technological advance in the infrastructure of mobile technologies. The increase in the use of wireless local area networks and the use of satellite services are also noticed. The high utilization rate of mobile devices for various purposes makes clear the need to track wireless networks to ensure the integrity and confidentiality of the information transmitted. Therefore, it is necessary to quickly and efficiently identify the normal and abnormal traffic of such networks, so that administrators can take action. This work aims to analyze classification techniques in relation to data from Wireless Networks, using some classes of anomalies pre-established according to some defined criteria of the MAC layer. For data analysis, WEKA Data Mining software (Waikato Environment for Knowledge Analysis) is used. The classification algorithms present a success rate in the classification of viable data, being indicated in the use of intrusion detection systems for wireless networks.

KEYWORDS

Wireless Networks, Classification Thecniques, Weka

1. INTRODUCTION

Over the past decade a great technological advance was seen, especially regarding mobile technologies and its infrastructure. The increase in the use of wireless local area networks and also the use of services from satellites, both in organizational and residential environments, is identified. This allows information to be created, transmitted and accessed faster and anywhere at any time by simply having access to the mobile network infrastructure. According to Anatel (Telecommunication National Agency), in January/2016 Brazil registered 257.248 million active lines in mobile telephony, with pre-paid accesses corresponding to 71.45% (183.80 million) of total accesses, while postpaid accesses correspond to 28.55% (73.45 million).

The consequence of this scenario is perceived when the use of computational devices used by both individuals and companies are verified. This scenario can be verified through the research conducted by IDC Brasil, which states that in the last quarter of 2014 Brazil had 1,637 million computers, of which 600 thousand are desktops and 1,037 million are notebooks. An unpublished survey by the Brazilian Institute of Geography and Statistics (IBGE) reveals that 57.3% of homes access the internet through cell phones and tablets in 2013.

People are getting used to technologies such as smartphones and tablets with Internet access. Most of these devices are equipped with capabilities based on the IEEE 802.11 standard. Using these wireless networks, users are often able to gain access to the Internet much cheaper than using cellular networks.

Currently these mobile devices basically act as a small computer, being possible to perform all actions, among others commonly performed on a Personal Computer. Some of these actions are: sending of E-mail to any computational device; use of an operating system; video viewing; execution of Web Systems; content servers; financial transactions; online shopping.

These mobile devices are also part of Wireless Networks as well as wireless actuators offering communication technologies for automation tools built into the Internet of Things in various environments [23].

The high rate of use of mobile devices for various purposes explains the importance of monitoring this infrastructure, since it presents the large-scale transmission of information, which at certain times may be restricted. To the set of this mobile system, determined by both the software and the hardware used, it is relatively fragile with regard to security, mainly due to the characteristic of its transmission medium, but also by the dynamism of access to this system. So there is a need to try to identify the normal and abnormal traffic of these wireless networks so that their administrators can take action.

With increased interconnection between networks, structured and wireless, information security has become a challenge. Networks are subject to various types of attacks that may have internal or external sources, some with the goal of paralyzing services, others with the intention of stealing information and in other cases, just for the amusement of the attackers. In addition, until recently, the networks were restricted to computers, now accept various types of equipment: sensors, smart phones, cell phones, among others. Therefore, security enhancement proposals should consider the technological evolution that is taking place.

The Wireless Networks environment, as well as the environment of Ad Hoc Wireless Networks or Wireless Sensor Networks, has in its characteristic a dynamicity in relation to the composition of the network members, that is, for these types of networks users often enter and leave the network. This feature makes it necessary to manage these environments quickly. This scenario becomes, however, quite vulnerable to attempts to approach the anomalies present in the system as a whole. Anomalies such as *EAPOL Start, Beacon Flood, Deauthentication, RTS Flood* [1][2].

However, the techniques and tools adopted by network managers in the framework of structured computer networks do not always meet these needs in a timely manner. In this sense, the use of intelligent algorithms for classification becomes a great option to minimize these difficulties, in order to identify anomalies more effectively.

The high rate of use of mobile devices for various purposes makes clear the need to monitor this infrastructure, since it presents the large-scale transmission of information, which at certain moments may be confidential. The set of this mobile system, determined by both the software and the hardware used, is relatively fragile regarding security, mainly due to the characteristic of its transmission mean, but also due to dynamic access it. So, there is a need to try to quickly and effectively identify the normal and abnormal traffic of these wireless networks so that administrators can take action. This work aims, from a database of wireless networks [1], to evaluate the classification of these data for some classification techniques. The data is formed by MAC layer information, which will be shown later.

The structure of this article is organized into sections. In section two will be presented some works that have the characteristic of identification of wireless networks traffic using algorithms of learning. In section 3, the theoretical basis for Wireless Networks is presented, while section 4 deals with Classification Techniques. Section 5 will present the methodology of experimentation and results. In Section 6 we present the case studies used to analyze the results.

In section 7 will be performed the quantitative and qualitative analysis of the results. Section 8 presents the conclusion of the work and future work.

2. RELATED WORKS

The large increase in the use of mobile computing resources both in public environments, both in private environment has aroused the great use of Ad Hoc Wireless Networks, mainly due to the ease of deployment of these networks. This, in turn, favors the large-scale development of malicious applications in Wireless Networks. It can be said that the number of attacks on Computer Networks, with wireless and structured architecture, has grown in recent years, with the incidents reported in the Brazil exceed 700,000, according to the Center for Studies and Responses to Security Incidents in Brazil [19]. Thus, there is a need to provide resources capable of guaranteeing the minimum authenticity of the services provided by the Computer Networks.

Intrusion Detection Systems are tools that contribute to guarantee the security in the Computer Networks, and its implementation is based on the policy of security of the environment with the objective of keeping active the services made available by the Computer Networks.

In addition, it is necessary to take into account the characteristics of the Ad Hoc Wireless Networks, which make it difficult to monitor the services and components of the Network, since they are constituted by autonomous nodes with mobility and without centralized management. Ad Hoc Wireless Networks rely on direct peer-to-peer communication, which is established without the need for centralized infrastructure. The Ad Hoc Networks are composed of devices that have the cooperative characteristic, being able to establish a direct communication with the devices that are within their reach. In this network there is centralized administration and each device can have the functionalities of station and router. The communication between the stations is called storage-forwarding, that is, the station that wishes to forward a message accesses the transmission medium and forwards the information to the neighboring station, which stores the information until the optimal time to forward the station other than the destination station. In this way, the formation of a multi-hop link between the information source and the destination of the information is identified, making network services such as routing and access control to the medium performed in a distributed way by all the components belonging to Ad Hoc Wireless Network [20].

There are several proposals of Intrusion Detection System in Wireless Networks [2, 5, 7, 21, 22] where the main obstacle is the durability of the energy of the computational resource, being frequently used in these technical proposals of computational intelligence capable of analyzing, learning and identifying anomalies. These proposals are based on the use of classification techniques, either in a single or joint approach, aiming increasingly to better use the mobile computing resource. The result of the use of classification techniques has contributed with the Computer Networks analysts in the choice of security policies with the purpose of nullifying or minimizing the damages caused by the anomalies in Wireless Networks environments.

It is possible to find in the literature some works of Wireless Networks traffic classification, which can be applied in Intrusion Detection Systems. These proposals make use of supervised and unsupervised learning methods. The proposal [2] provides a general approach to the various classification methods, using high-dimensional data and a variable selection technique aiming to reduce computational time and improving the learning rate.

Govindarajan presents a proposal [3] of two classification methods involving multilayer perceptron and Basis function Networks. This work proposes a hybrid architecture involving both classifiers for intrusion detection systems. Ed Wilson presents a proposal [4] of Hybrid

Intrusion Detection System, in which signal processing is performed using the Wavelet transform and then the classification of the anomalies using Artificial Neural Networks.

Ed Wilson[1] proposes the elaboration of a real database of Wireless Network traffic, which will be used in the evaluation of Intrusion Detection Systems (IDS). This data undergoes a pre-processing to later be classified by techniques of standards recognition, such as Artificial Neural Networks and following formatting rules that must be strictly followed.

The proposal [5] uses a combination of selection methods to classify Denial Of Service anomalies in Computer Networks, showing the efficiency of the process selection process for DoS detection.

Vo [6] applies supervised and unsupervised machine learning techniques to predict the time series trend by using the K-Means algorithm to group data with similarity and vector machine to train and test the data.

In [7] the most relevant models for the construction of Intrusion Detection Systems are presented, incorporating machine learning in the scenario of Ad Hoc Wireless Networks. Machine learning methods perform classification approach, association rule mining, Artificial Neural Networks and instance-based learning.

Work [21] also uses unsupervised and supervised classification methods to classify a collection of packet data from the Internet.

Gogoi presents the proposal [22] of a multi-level hybrid intrusion detection method that uses a combination of supervised, unsupervised and discrepant-based methods to improve the efficiency of detecting new and old attacks.

3. WIRELESS NETWORKS

The IEEE 802.11 standard defines a structure for the Wireless Local Area Network that covers the physical and link levels present in the reference OSI communication model. For the physical level only, radio frequency (RF) and infrared (IR) transmissions are treated, but other forms of wireless communication such as microwave and visible light can also be considered. For the link level, the access control to the medium is addressed through the definition of the MAC protocol (Medium access Control).

Taking into account the main characteristics of the IEEE 802.11 standard, such as interoperability, low cost, high market demand, reliability of project execution, there is a great growth in the use of Local Area Networks of Wireless Computers, also known as Wireless Networks, in public and private environments. This makes Wireless Networks a priority resource in environments where it is most often possible to access the Internet, whether inside corporations, in homes or in public environments, such as shopping malls, airports and so on [1].

The architecture of Wireless Networks according to the IEEE 802.11 standard is based on the division of the area covered by the Wireless Network into cells, these cells being called BSA (Basic Service Area). The size of the coverage of each BSA will depend exclusively on the characteristics of the environment itself and the power of transmitters and receivers used in the computational devices. The other components of the Wireless Networks architecture are listed below[1]:

I. BSS (*Basic Service Set*): Which is the set of computational devices that communicate by broadcasting (BC) or infrared (IR) within a Basic Service Area;

II. AP (*Access Point*): Specific computational devices, which have the purpose of capturing the transmissions made by computational devices belonging to its BSA (Basic Service Area) and are destined to stations belonging to another Basic Service Area. The Access Point, in turn, will perform the retransmission using a distribution system;

III. Distribution System: Communication infrastructure, which has the purpose of performing the interconnection of several Basic Service Area to allow the construction of networks, which have covers larger than one cell;

IV. ESA (*Extended Service Area*): Service Area that has the purpose of interconnecting several BSAs, through the Distribution System using the Access Point;

V. ESS (*Extended Service Set*): Which is intended to represent a set of computational devices consisting of the union of several BSSs (Basic Service Set) connected by a Distribution System.

The IEEE 802.11 standard also defines a medium access protocol, which is present in a MAC sublayer of the data link level. This protocol is called DFWMAC (*Distributed Foundation Wireless Medium Access Control*), which has two access methods, one of which is a distributed and mandatory feature. The other access method of the DFWMAC protocol is optional, having a centralized feature, and according to the IEEE standard, both the distributed method and the centralized method in the communication system can coexist. The medium access protocol also has the property of treating problems related to computational devices that try to move from one cell to another, a process called roaming. It is also related to the protocol of access to the medium of property to treat problems of lost computational devices, being able to be denominated of hidden node [1].

4. CLASSIFICATION TECHNIQUES

Classification is one of the Data Mining techniques that is mainly used to analyze a given dataset and takes each instance of it and assigns this instance to a particular class, thus granting a low error of classification. It is used to extract models that accurately define important data classes within the given dataset. Classification is a two-step process. During first step the model is created by applying classification algorithm on training data set then in second step the extracted model is tested against a predefined test dataset to measure the model trained performance and accuracy. So classification is the process to assign class label from dataset whose class label is unknown.

The dataset evaluation relied on the following classifiers: Bayesian networks, decision tables, Ibk, J48, MLP and NaiveBayes. The main criteria used were the popularity of such classifiers.

Bayesian networks have been used in many approaches to IDS, as in UMER (2017) [8]. These networks are directed acyclic graphics for representing a probability distribution on a set of random variables. Each vertex represents as random variable and each node represents a correlation among the variables [1] [9].

The decision table classifier works representing a set of conditions needed to determine the occurrence of a group of actions by means of a table format [10]. This technique has also been used in IDS approaches [1][11].

The IBk algorithm refers to a way of implementing the kNN (k-nearest neighbor) clustering method, which is used for classification and regression toward finding the closest neighbors of a given instance. In the IBk, three neighbors, the ones closest to the search standard neighbors, are used. This is a relatively simple technique that has been used in IDS approaches as well [1][12].

The J48 algorithm relies on decision tree classifications. By this technique, the classification of a new item depends on the prior creation of a decision tree which uses attributes obtained from the training data. By computing the information gain of each of these attributes, J48 can optimize classification mechanisms in IDS [1][13].

The MLP is an artificial neural network that maps input parameters to proper outputs. It consists of many layers of nodes in a directed graphic. Several IDS approaches have used MLP [1][14].

5. METHODOLOGY

This work aims to apply classification techniques to identify anomalies especially in wireless network traffic. As mentioned in the previous section, the techniques adopted for this work are: Bayesian Networks, Decision Tables, Ibk, J48, MLP and NaiveBayes.

In order to achieve the proposed aims, the following activities were performed in accordance with the chronological order of execution.

We use a database with examples of specific anomalies in wireless networks. This base in turn is the final product of the work entitled *A Methodology for building a Dataset to Assess Intrusion Detection Systems in Wireless Networks* [1].

The next step is to perform a pre-processing in the database so that two new databases are obtained. One of the databases is composed of only 10% of the data from the original database and is destined for the test step in the selected algorithms. The other database is composed of 90% of the data from the original database and is destined for the training step of the selected algorithms. Both databases are stored in the Database Manager System named PostgreSQL, and are accessed by the Weka software (*Waikato Environment for Knowledge Analysis*)[15].

Finally, the results of each selected algorithm are analyzed and formatted through tables. In relation to the results, the following information is presented for analysis: Percentage of Classification, relation of correctness and errors.

6. CASE STUDY

The case study chosen to analyze the results of the application of classification techniques presented in previous sections uses data from real wireless networks [1] and the data mining software, Weka [15].

6.1. Database

The database defined for the execution of this case study is a real collection of network traffic captured in the Wireless architecture. This data, in turn, is obtained by the behavior of users to access different information as well as for the use of the Internet. According to the authors [1], the network traffic obtained by students and employees of the institution in which the experiment was performed was used for this database.

The database chosen for the experimentation of this work made use of two different scenarios. The scenarios discussed have their own configuration and topologies, being a scenario of home environment typical of wireless networks, while the other is a more complex environment, being a corporate environment.

This database is composed of a total of 616,047 records, each record being composed of 16 variables that are characteristics of the wireless network traffic itself. Also in each record of the database is defined a last variable the class to which belongs certain registry, classification is realized taking into account the values of the sixteen variables referring to the obtained wireless network traffic. In this way the data are classified in:

- *Normal*: Acceptable wireless network traffic;
- *EAPOLStart*: Traffic using the Extensible Authentication Protocol (EAP), which aims to perform an authentication method in both the Wired Equivalent Providence (WEP) protocol, both Wi-Fi Protected Access (WPA) protocol, commercial versions for wireless network access;
- *Beacon Flood*: Management type requests, which are intended to transmit millions of invalid Beacons, resulting in the difficulty that a certain Wireless network device will have in identifying a legitimate Access Point [16];
- *Deauthentication*: It also represents management-type requests, which are injected from the Wireless Network. The frames belonging to this anomaly are transmitted as fictitious requests, which requests the deactivation of a device that is authorized in the Wireless Network;
- *RTSFlood*: Also called Request-to-Send Flood is a control-type frame. This anomaly is based on the large-scale transmission of RTS frames or frames for a short period of time [16].

The database for the experimentation process of this work is divided into two distinct bases, in order to meet the requirements of each defined intelligent algorithm. In this way a training database is generated respecting the characteristics of each algorithm, being composed by 554,442 registers, which corresponds to 90% of the complete database. Also, the test database is generated, being composed by 61,604 records that correspond to 10% of the complete database, respecting the characteristic of each algorithm. In order to optimize the experimentation process and to provide better data manipulation, the training and test databases for each defined computational intelligence technique are stored in the PostgreSQL Database Management System.

6.2. Dataset Evaluation

The data coming from Wireless Network are evaluated through the classification techniques mentioned in the previous section. To evaluate each of the classification techniques, the error parameters, the percentage of classification and the Kappa coefficient are used, which will be explained later.

The Mean Absolute Error (MAE) is defined as the average of the difference between and computed and measured results. The closer to zero the better the classification is. On the other hand, the Root Mean Square Error (RMSE) is computed as the average of the error square root. A minimum MAE does not imply necessarily in a minimal variation. Thus, it is more effective to use both MAE and RMSE in the evaluations [17].

The MAE and RMSE parameters are a simple way of measuring the effectiveness and efficiency of the classification techniques used, thus they are incentive of more advanced techniques.

The Kappa coefficient, in turn, is initially used by observers in the field of psychology as a measure of agreement-induced [18]. This metric shows the degree of acceptance or agreement among a group of judges. Equation 1 shows the agreement of the Kappa coefficient, with the observed agreement *Po* and the coincidence by chance *Pa*.

$$k = \frac{Po - Pa}{1 - Pa} \tag{1}$$

The result of $k = 1$ means that the classification was correct, while $k = 0$ indicates that the classification is entirely by chance. However, the best classifiers are those in which the value of k is close to one.

As previously shown, the classification techniques to be evaluated for the Wireless Networks database are: Bayesian networks, decision tables, Ibk, J48, MLP and NaiveBayes.

6.2. Results and Discussion

The evaluation of the classifiers is performed using the Weka [15] tool, using the set of data obtained from Wireless Network [1] in which they are classified with the following anomalies: EAPOLStart, Beacon Flood, Deauthentication and RTSFlood. This database is composed of 17 variables per record, 16 MAC layer attributes and an identification attribute of the class to which a particular record belongs.

Experimentation with the chosen classification techniques makes use of 90% of training data and 10% of data for testing. The results of the mean and quadratic errors for the same, during the training are shown in Figure 1. These are relatively small, and it can be deduced that the classifiers have good performance for the data set of Wireless Networks.

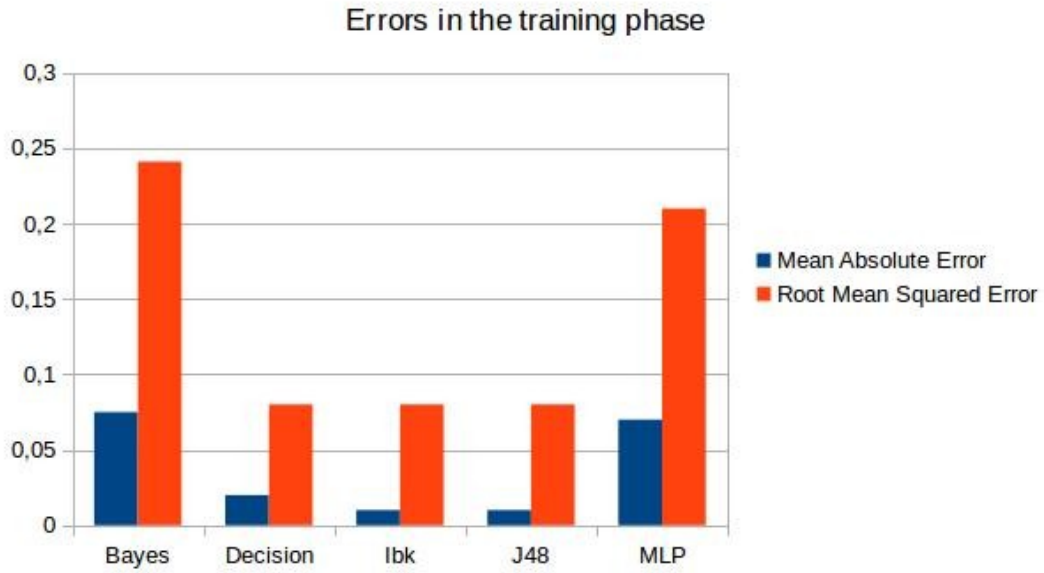


Figure 1. Errors in the training phase

Table 1 presents the simulation results, after the training of the classification techniques in relation to the Wireless Network data. The values obtained in percentage of correctly classified instances are relevant, being superior to some found in literacy. It should be considered that the proposal is to evaluate the performance of the classification techniques for the application of Wireless Networks data, without customizing them.

Table 1.Results for the testing phase of the data set

Classification Techniques	**Correctly Classified Instances (%)**	**Incorrectly Classified Instances (%)**	**Kappa Coefficient**
Bayes Network	76	24	0,42
Decision Tree	98	2	0,91
Ibk	98	2	0,91
J48	98	2	0,91
MLP	75	25	0,4

The evaluation of the data set represents an important research phase in the area of Wireless Networks, as it allows verifying the adequate response of the classification techniques commonly used in Intrusion Detection Systems proposals.

The use of the classification techniques adopted showed good results. The average errors, as shown in Figure 1 are relatively low. It is observed that the absolute mean error as well as the mean square error followed the same trend, proving the actual behavior of the data of Wireless Networks.

Table 1 shows that there is no difference for similar classification algorithms such as Bayes Network and MLP, in which it obtained a rating of 75%, while the other classification algorithms reached 98% of classification with low average errors. Therefore, it is possible to affirm that the use of classification techniques are effective for Wireless Network environments and can be used in Detection and Anomaly Classification Systems for Wireless Networks. It is also noticed that the selection of variables is fundamental for the classification to reach satisfactory levels and optimize the processing of these algorithms.

7. CONCLUSIONS AND FUTURE WORK

The results show that the data used to evaluate the classification techniques are viable and can be components in the evaluation of Intrusion Detection Systems in Wireless Networks. However, despite being preformatted with labels, where each record is identified as normal or with some of the predefined anomalies, it becomes valuable because it is collected directly from a Wireless Network.

The errors found in the training phase of the classification algorithms are low, being below 0.25, confirming that the selected classification techniques are adequate and that the data collected from Ad Hoc Wireless Networks are efficient for the analysis of the same ones.

The Kappa Coefficient results follow the same characteristics of the errors in the training phase of the classification algorithms in relation to the correct and incorrectly classified data, thus confirming their integrity.

Future work can be done in several ways: applying these classification techniques to a wireless network, online detection and classification on the network, and comparing with other existing approaches.

REFERENCES

[1] E. W. T. Ferreira, et al., (2015) “A methodology for building a dataset to assess intrusion detection systems in wireless networks,” WSEAS Transactions on Communications, vol.14, pp. 113–120, 2015.

[2] G. C. F. Sahin, (2014) “A survey on feature selection methods,” Computers Electrical Engineering, 2014, pp. 16–28.

[3] M. Govindarajan & R. M. Chandrasekaran, “Intrusion detection using neural based hybrid classification methods,” Computer Networks vol. 55, no. 8, pp. 1662–1671, 2011.

[4] E. W. T. Ferreira, et al., (2011) “Intrusion detection system with wavelet and neural artifical network approach for networks computers,” IEEE Latin America Transactions, vol. 9, no.5, pp. 832–837, 2011.

[5] S. Bhattacharya & S. Selvakumar, (2016) “Multi-measure multi-weight ranking approach for the identification of the network features for the detection of dos and probe attacks,” The Computer Journal, vol. 59, no. 6, pp. 923–943, 2016. [Online]. Available: <http://dx.doi.org/10.1093/comjnl/bxv078>

[6] V. Vo & J. Luo & B. Vo, (2016) “Time series trend analysis based on k-means and support vector machine,” Computing and Informatics, vol. 35, p. 11–127, 2016. [Online]. Available: <https://pdfs.semanticscholar.org/fd6d/6d3778f52608f048aa95dd9aaca42fe2871f.pdf>

[7] L. Nishani & M. Biba, “Machine learning for intrusion detection in manet: a state-of-the-art survey,” Journal of Intelligent Information Systems, vol. 46, no. 2, pp. 391–407, 2016.

[8] UMER, Muhammad Fahad & SHER, Muhammad & BI, Yaxin. (2017) “Flow-based intrusion detection: Techniques and challenges”. **Computers & Security**, v. 70, p. 238-254, 2017.

[9] N. Friedman & D. Geiger & M. Goldszmidt,(1997) “Bayesian network classifiers,” Mach. Learn., vol. 29, no. 2–3, pp. 131–163, 1997.

[10] Wei, W. & Wang, J. & Liang, J. & Mi, X. & Dang, C. (2015). Compacted decision tables based attribute reduction. *Knowledge-Based Systems*, *86*, 261-277.

[11] Rajmahanty, P. H., & Ganapathy, S. (2017). Role of Decision Trees in Intrusion Detection Systems: A Survey. *International Journal of Advances in Computer and Electronics Engineering*, *2*(4), 09-13.

[12] Modi, M. U., & Jain, A. (2015). A survey of IDS classification using KDD CUP 99 dataset in WEKA. *Int. J. Sci. Eng. Res*, *6*(11), 947-954.

[13] Aljawarneh, S., Yassein, M. B., & Aljundi, M. (2017). An enhanced J48 classification algorithm for the anomaly intrusion detection systems. *Cluster Computing*, 1-17.

[14] S. Haykin, Neural Networks and Learning Machines, 3nd.Ontario Canada: Pearson, 2009.

[15] Modi, M. U., & Jain, A. (2015). A survey of IDS classification using KDD CUP 99 dataset in WEKA. *Int. J. Sci. Eng. Res*, *6*(11), 947-954.

[16] R. F. de Moraes & N. V. D. S. A. C. Maciel, “Avaliação de um conjunto de dados quanto á sua qualidade na especificação de perfis de ataque e não-ataque numa rede IEEE 802.11w,” Anais da VI Escola Regional de Informática da Sociedade Brasileira da Computação(SBC) - Regional de Mato Grosso, 2015, pp. 145–1508.

[17] Terziyska, M. & Todorov, Y. & Dobreva, M. (2018). Efficient Error Based Metrics for Fuzzy-Neural Network Performance Evaluation. In *Advanced Computing in Industrial Mathematics* (pp. 185-201). Springer, Cham.

[18] J. Cohen, “A Coefficient of Agreement for Nominal Scales,” Educ. Psychol. Meas., vol. 20, no. 1, pp. 37–46, Apr. 1960.

[19] CERT-BR, “Estatísticas dos incidentes reportados ao cert.br,” <https://www.cert.br/stats/incidentes/>, accessed: 2018-06-01.

[20] J. Loo, J. L. Mauri, and J. H. Ortiz, Mobile ad hoc networks: current status and future trends. Press, 2016.

[21] A. Vlăduţu, D. Comăneci, and C. Dobre, “Internet traffic classification based on flows’ statistical properties with machine learning,” International Journal of Network Management, vol. 27, no. 3, p.1929, 2017.

[22] P. Gogoi, D. Bhattacharyya, B. Borah, and J. K. Kalita, “Mlh-ids: A multi-level hybrid intrusion detection method,” The Computer Journal, vol. 57, no. 4, pp. 602–623, 2014. [Online]. Available: <http://dx.doi.org/10.1093/comjnl/bxt044>

[23] M. Sha, D. Gunatilaka, C. Wu, and C. Lu, “Empirical study and enhancements of industrial wireless sensor–actuator network protocols,” IEEE Internet of Things Journal, vol. 4, no. 3, pp. 696–704, 2017.

Authors

Daniel R. Canêdo has a degree in Computer Engineering from Pontifícia Universidade Católica de Goiás (2003) and a Master's degree in Electrical Engineering from the University of Brasília (2006). He is currently an exclusive professor at Federal Institute of Goiás - Campus Luziânia. He is currently a PhD student in the Post-Graduate Program in Electronic Systems and Automation Engineering of the Department of Electrical Engineering of the University of Brasília (UnB).

Alexandre R. Romariz holds a BS in Electrical Engineering from the University of Brasília (1992), a Master's degree in Electrical Engineering from the State University of Campinas (1995) and a PhD in Electrical Engineering from the University of Colorado at Boulder (2003). He is currently “Professor Associado” at University of Brasilia. He has experience in Computational Intelligence, Integrated Circuits, Optoelectronics and Digital Signal Processing.

C. APÊNDICE(ARTIGO SUBMETIDO - IEEE LAT) - INTRUSION DETECTION SYSTEM IN AD HOC NETWORKS WITH NEURAL NETWORKS ARTIFICIAL AND K-MEANS ALGORITHM

Intrusion Detection System in Ad Hoc Networks with Neural Networks Artificial and K-Means Algorithm

D. R. Canêdo and A. R. S. Romariz

***Resumo*—There has been a great technological advance in the infrastructure of mobile technologies. The increase in the use of wireless local area networks and the use of services from satellites is also noticeable. The high rate of use of mobile devices for various purposes brings the need to monitor the wireless networks to ensure the integrity and confidentiality of the information. Therefore it is necessary to quickly and efficiently identify the normal and abnormal traffic of these wireless networks so that their administrators can take action. This paper presents a proposal for a Ad Hoc Wireless Intrusion Detection System composed of two stages, based on data grouping through the algorithm K-Means and Artificial Neural Networks through the Multilayer Perceptron algorithm, for the detection and classification of anomalies caused by attacks on the networks of Computers.**

***Index Terms*—Ad Hoc Wireless Networks, Multilayer Perceptron, K-Means, Intrusion Detection System.**

I. Introdução

REDES de Computadores tem visto um aumento significativo em sua infraestrutura de conexão, tornando a segurança da informação um desafio. Atualmente, há um aumento no uso de Redes de Computadores Sem Fio, tanto em ambientes residenciais quanto no ambiente corporativo. Dados recentes da Anatel (Agência Internacional de Telecomunicações) mostram que cerca de 260 milhões de brasileiros usam dispositivos móveis para acessar a Internet para criar, transmitir ou consumir informações [2].

Redes Ad Hoc segue o padrão 802.11, sendo definida como Redes de Computadores Sem Fio sem a presença de um componente concentrador, tornando cada nó da Rede responsável pelo roteamento e controle de acesso ao meio e gerenciando algumas características da Rede como: Baixa taxa de transmissão; Probabilidade de erro; Variações no meio de transmissão. Essas redes são formadas em ambientes onde há necessidade de comunicação, mas há uma inoperabilidade de Redes Sem Fio com estrutura, tornando as Redes Ad Hoc de natureza temporária e complexa [13].

No entanto, Redes Ad Hoc estão sujeitas a ataques que podem ter origem interna e externa, alguns deles com o objetivo de paralisar alguns serviços dos nós da própria Rede, enquanto outros têm o objetivo de capturar informações que trafegam entre os nós de Redes Ad Hoc.

D. R. Canêdo, Universidade de Brasília, Brasília, Distrito Federal, Brasil e Instituto Federal de Goiás, Luziânia, Goiás, Brasil, daniel.canedo@ifg.edu.br.

A. R. S. Romariz, Universidade de Brasília, Brasília, Distrito Federal, Brasil, alromariz@gmail.com.

A confidencialidade, integridade e disponibilidade dos recursos das Redes são fundamentais para prover a segurança da informação, sendo que um processo de anomalia em Redes de Computadores incluindo as Redes Ad Hoc podem comprometer sistemas, caracterizando uma intrusão. O IDS (*Intrusion Detection System*) tem o propósito de identificar intrusões em Redes de Computadores, sem comprometer o funcionamento normal da Rede. O Sistema de Detecção de Intrusão é considerado uma ferramenta de segurança de Redes, que em conjunto com outras ferramentas de segurança são organizadas para reforçar a segurança da informação em sistemas de comunicação [1].

A análise do tráfego de rede nas Redes AdHoc é dificultada pela falta de gerenciamento central. Outra característica importante a considerar é a alta mobilidade dos componentes da Rede Ad Hoc, já que pode-se entrar e sair da Rede sem restrições. Outra característica é que os componentes da Rede Ad Hoc são na maioria das vezes dispositivos móveis, que possuem restrições em seu estado ativo, pois dependem da energia de seus recursos. Estas características das Redes Ad Hoc remetem que os Sistemas de Detecção de Intrusão tradicionais não são usados diretamente.

Este artigo apresenta uma proposta de Sistema de Detecção de Intrusão para Redes Ad Hoc através de duas etapas. A primeira etapa destina-se a agrupar todos os tráfegos da Rede de um determinado nó através da utilização do Algoritmo K-Médias, enquanto o segundo passo é classificar as anomalias levando em conta as informações do grupo, através da aplicação de Redes Neurais Artificiais com o algoritmo *Multilayer Perceptron*.

A estrutura deste artigo esta organizada em seções. Na seção dois serão apresentadas propostas para Sistemas de Detecção de Intrusão. Na seção 3 apresenta-se a fundamentação teórica abordando Redes Sem Fio Ad Hoc, enquanto que na seção 4 aborda-se Sistemas de Detecção de Intrusão em Redes Ad Hoc. Na seção 5 será apresentada a abordagem proposta do Sistema de Detecção de Intrusão, bem como os resultados da simulação da mesma. Na seção 6 apresenta-se a conclusão do trabalho e a apresentação de trabalhos futuros.

II. Trabalhos Relacionados

Na literatura existem trabalhos de classificação de tráfego de Redes Wireless, os quais podem ser aplicados em Sistemas de Detecção de Intrusão. Estas propostas utilizam métodos de aprendizagem supervisionados e não supervisionados.

A proposta de Chandrashekar(2014) [4] fornece uma abordagem geral dos vários métodos de classificação, usando dados de alta dimensão e uma técnica de seleção de variáveis com o objetivo de reduzir o tempo computacional e a velocidade de aprendizagem.

Govindarajan apresenta uma proposta [10] de dois métodos de classificação envolvendo perceptron multicamada e função de base radial. Propõe-se neste trabalho uma arquitetura híbridaa envolvendo ambos os classificadores para sistemas de detecção de intrusão.

Cervantes apresenta uma proposta [3] de sistema de detecção de intrusão contra ataques *sinkhole* e *selective forwarding* sobre o roteamento na IoT densa e móvel. Utiliza agrupamento para lidar com a densidade e a mobilidade, e combina estratégias de *watchdog*, reputação e confiança na detecção de atacantes, a fim de garantir confiabilidade aos dispositivos.

EdWilson apresenta uma proposta [8] de Sistema de Detecção de Intrusão híbrido, em que realiza-se um processamento de sinais através da utilização de transformações Wavelets e posteriormente a classificação das anomalias utilizando Redes Neurais Artificais.

EdWilson apresenta uma proposta [9] que propõe a elaboração de uma base de dados reais de tráfego de Redes Wireless, a qual será utilizada na avaliação do Sistema de Detecção de Intrusão - IDS - proposto. Estes dados por sua vez sofrem um pré-processamento para posteriormente serem classificados por técnicas de reconhecimento de padrões, como por exemplo Redes Neurais Artificiais.

III. Redes Sem Fio Ad Hoc

Uma Rede Ad Hoc é formada em situações onde há necessidade de comunicação e uma infraestrutura fixa não está disponível ou não é desejável [6]. Nesse caso, os nós móveis formam uma rede para uso temporário, a fim de atender às necessidades de comunicação naquele momento, ou ad hoc. Uma Rede Sem Fio Ad Hoc, também denominada de MANET, é um sistema de rede sem fio com nós móveis que podem mover-se livremente e são auto-organizáveis com topologia dinâmica e permite que equipamentos possam utilizar a rede sem comunicação preexistente, diferente de uma rede com infraestrutura fixa [6].

A Figura 1 [7] apresenta um modelo de Rede Sem Fio Ad Hoc que permite a comunicação diretamente entre os nós, e estes por sua vez podem realizar o repasse de pacotes através de múltiplos saltos. Cada elemento da rede é responsável pelo encaminhamento de pacotes de seus vizinhos. Cada nó é equipado com uma ou mais interface de rádio, e a cobertura da rede depende diretamente do alcance destes enlaces. Até certo ponto, é possível adicionar mais nós na rede e, consequentemente, aumentar sua cobertura.

Os nós da rede também podem funcionar como roteadores para outros nós, com o encaminhamento de pacotes para o destinatário final. Esta rede pode possuir conexão com rede com infraestrutura, através de *gateways*. Alguns exemplos de aplicações deste modelo de rede são campos de batalhas militares, locais onde existe necessidade de formação rápida de redes, missões de resgate, redes de sensores para automação e aplicações em eventos [15].

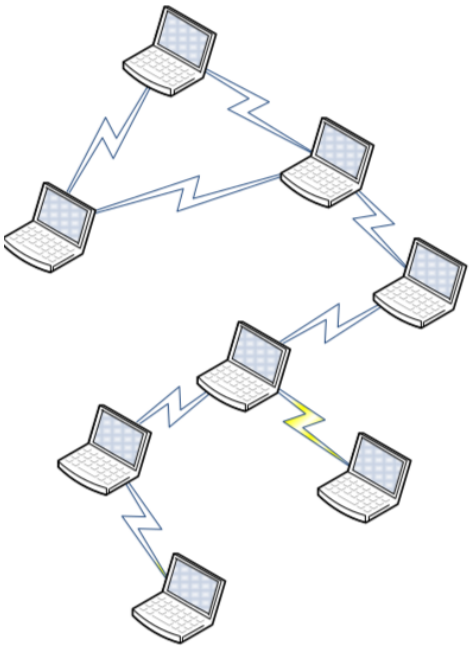


Figura 1. Redes Sem Fio Ad Hoc[7]

IV. Sistema de Detecção de Intrusão em Redes Ad Hoc

Uma intrusão é definida como certas ações cuja finalidade é comprometer as propriedades de confidencialidade, integridade e disponibilidade dos recursos da Rede de Computadores. Um Sistema de Detecção de Intrusão - IDS - deve ser capaz de identificar ações maliciosas, no entanto, não deve comprometer a operação da Rede de Computadores. O IDS, por outro lado, deve consumir poucos recursos computacionais, para não prejudicar usuários legítimos.

Confidencialidade, integridade e disponibilidade de recursos representam fatores vitais para a segurança da informação, onde uma ação maléfica ou não intencional pode comprometer o sistema, caracterizando uma intrusão. O sistema de detecção deve conseguir identificar essa ação, mas sem comprometer o funcionamento normal da rede. Um sistema de detecção é uma ferramenta de segurança que, como outras medidas, a exemplo de antivírus e *firewalls*, destinam-se a reforçar a segurança da informação em sistemas de comunicação [1].

Os IDS's são usados para monitorar, avaliar e informar violações de segurança que podem ser intencionais ou não. No entanto, as técnicas de detecção e prevenção não avançam no mesmo ritmo, o que dificulta sua aproximação.

De uma forma geral os Sistemas de Detecção de Intrusão tradicionais não são empregados diretamente nas Redes Ad Hoc devido à particularidade de sua infraestrutura, pois apresenta influência direta no funcionamento do IDS. Atualmente existem propostas envolvendo Sistemas de Detecção de Intrusão em Redes Ad Hoc, sendo algums delas apresentadas na seção de Trabalhos Relacionados.

V. Abordagem Proposta

A proposta deste artigo caracteriza-se por um Sistema de Detecção de Intrusão em Redes Ad Hoc através de duas etapas, sendo a primeira dedicada ao agrupamento dos dados em grupos provenientes de uma Rede Sem Fio Ad Hoc, enquanto a segunda etapa é responsável pela classificação de anomalias pré-definidas.

O Sistema de Detecção de Intrusão proposto faz uso das Técnicas de Inteligência Computacional para executar as duas etapas relatadas. Para a primeira etapa, o agrupamento de dados do tráfego da Rede Ad Hoc é realizado pelo algoritmo K-Médias, que é um algoritmo de aprendizado não supervisionado. O algoritmo K-Médias separa certos objetos em grupos, chamados *clusters*. Estes *clusters* são formados através da aplicação de técnicas de medição de distância ou técnicas de similaridade entre objetos [14]. Este algoritmo é escolhido principalmente pela simplicidade computacional. O algoritmo K-Médias é capaz de processar grandes volumes de dados, cuja complexidade de armazenamento é O((m + K)n), onde *m* é o número de pontos e *n* é o número de atributos [14].

Após o agrupamento de dados da Rede Ad Hoc, o segundo passo é realizado para classificar anomalias pré-definidas utilizando a técnica de inteligência computacional Redes Neurais Artificiais através do algoritmo *Multilayer Perceptron*. O algoritmo *Multilayer Perceptron* será composto por pelo menos uma camada oculta entre a entrada e a saída. As camadas ocultas não possuem conexões com o mundo exterior, como mostra a Figura 2. Esse tipo de Rede Neural está sendo utilizado em larga escala para resolver problemas complexos, pois tem como característica o treinamento supervisionado com o processo de Correção de Erros, como o algoritmo de retropropagação [12]. A Figura 3 mostra o comportamento de um neurônio no processo de Aprendizagem por Correção de Erros, com os elementos fundamentais sendo o vetor de entrada, camada de neurônio oculto, neurônio de saída, função de soma.

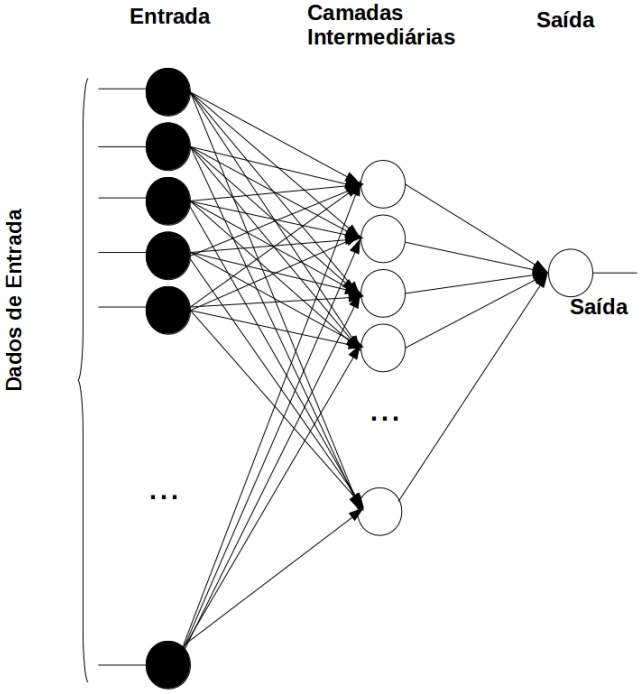


Figura 2. Redes Neurais de Várias Camadas

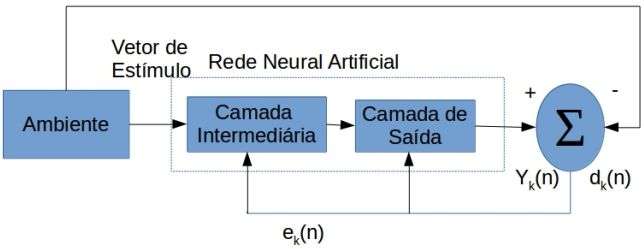


Figura 3. Aprendizagem por Correção de Erro

A. Base de Dados

A base de dados utilizada neste trabalho é uma coleção real de tráfegos de rede capturados na arquitetura Ad Hoc. Estes dados por sua vez representam o comportamento de usuários que frequentemente utilizam a Rede Sem Fio Ad Hoc para acessar diversas informações, bem como para a utilização da Internet. Segundo Ferreira(2015) [9] a base de dados é construída a partir do tráfego obtido pela comunidade acadêmica da instituição na qual o experimento é realizado.

Para a coleta dos dados utiliza-se dois cenários distintos, pois tem o objetivo de aumentar as possibilidades de tráfego da Rede. Os cenários abordados possuem configurações e topologias próprias, sendo um cenário representando um ambiente doméstico típico de Redes Wireless, enquanto o outro cenário é um ambiente um pouco mais complexo, corporativo.

Esta base de dados é composta por um total de 616047 registros, sendo que cada registro é composto por 16 variáveis que são características do próprio tráfego de rede. Além disso, em cada registro do banco de dados, a classe à qual pertence determinado registro é definida. A classificação é realizada nas seguintes classes:

- Normal: Dados que possuem características de tráfego de Redes Wireless aceitáveis;
- *EAPOLStart*: Uso do protocolo *Extensible Authentication Protocol*(EAP), cujo o objetivo é realizar um método de autenticação tanto na utilização do protocolo *Wired Equivalent Provacy*(WEP), tanto para o protocolo *Wi-Fi Protected Access*(WPA), em suas versões comerciais para acesso a Rede Sem Fio. Esta anomalia se caracteriza por uma carga excessiva de solicitação EAPOL - Start, que em um sobrecarregamento do *Access Point*, responsável pela interconexão dos dispositivos da Rede Wireless;
- *BeaconFlood*: Solicitações de tipo de gerenciamento, destinadas a transmitir milhões de Beacons inválidos, dificultando que um componente sem fio específico identifique um ponto de acesso legítimo. Esses *beacons* ajudam a identificar a localização do BSS (*Basic Set Service*) de uma Rede Wireless [5];
- *Deauthentication*: Solicitações do tipo gerenciamento,

que são injetados na Rede Wireless. Os quadros pertencentes a esta anomalia são transmitidos como pedidos imaginários, os quais solicitão a desautenticação de um dispositivo que se encontra autorizado na Rede Wireless;
- *RTSFlood*: Denominado *Request-to-Send Flood* é um quadro do tipo controle. Esta anomalia se baseia na transmissão em grande escala de pacotes ou frames RTS por um curto perıodo de tempo. A inundação de frames RTS na Rede Wireless proporcionará o congestionamento na reserva do canal Wireless, resultando no processo de negação de serviço aos nós da Rede Wireless [5].

B. Modelo Proposto

Nesse trabalho é proposto o uso de um Sistema de Detecção de Intrusão local com detecção em duas etapas. A primeira agrupa os dados da Rede, enquanto que a segunda, classifica os ataques. A primeira etapa analisa os dados e gera *clusters*, que são compostos por dados que possuem uma similaridade.

A segunda etapa é formada por uma Rede Neural Artificial, com cinco neurônios na saıda. Essa rede é treinada para reconhecer cinco classes, sendo quatro classes de ataques e uma classe de tráfego normal, como apresentado anteriormente.

A Rede Neural é formada por um MLP (*Multilayer Perceptron*) treinado com o algoritmo *backpropagation* [11]. A entrada é composta por 17 neurônios, sendo 16 deles referentes às variáveis da Rede Sem Fio Ad Hoc e 1 neurônio referente a informação do *cluster* resultante da etapa anterior. A camada oculta é formada por 10 neurônios.

O algoritmo para o Sistema de Detecção de Intrusão proposto, de forma simplificada é apresentado na Figura 4. Os dados da Rede, são obtidos através da captura do tráfego da rede. Note que o algoritmo fica continuamente executando para analisar os dados e gerar seus *clusters*, através do algoritmo K-Médias. Em seguida, o algoritmo *Multilayer Perceptron* classifica as anomalias reconhecidas para posterior comunicação com o gestor da Rede.

A proposta deste Sistema de Detecção de Intrusão, conforme apresentado ns Figura 4, deve ficar em execução continuamente, para obter os dados da rede. É realizado um pré-processamento dos dados que contenham características importantes da arquitetura de Redes Sem Fio Ad Hoc, como por exemplo quadros da camada de enlace, a exemplo dos quadros de solicitação de associação em pontos de acesso. Com isso, será possível identificar anomalias exclusivas dessa arquitetura. Como os dados são obtidos diretamente na rede, não é esperado aumento de *overhead*, independente da arquitetura utilizada.

O próximo passo após realizar o pré-processamento dos dados obtidos da Rede Sem Fio Ad Hoc é a geração dos *clusters* através da execução do algoritmo K-Médias. Para a execução do algoritmo K-Médias define-se a construção de 10 *clusters* e a utilização da função de distância Euclidiana para medir a similaridade entre os dados de cada grupo.

O terceiro passo do algoritmo proposto é classificar as anomalias existentes na base de teste, através da execução do algoritmo *Multilayer Perceptron*. Para a execução do algoritmo *Multilayer Perceptron* é definido uma Rede Neural com 17

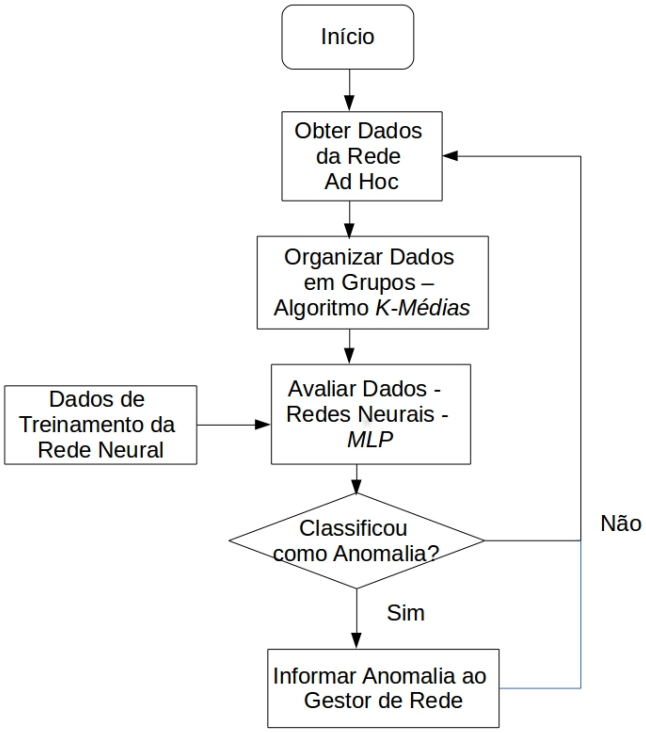


Figura 4. Algoritmo Proposto

neurônios na entrada(16 referentes à rede sem fio Ad Hoc e 1 referente ao *cluster*) e 10 neurônios para a camada oculta, como mostra a Figura 5. O algoritmo *Multilayer Perceptron* de acordo com o algoritmo proposto é treinado com dados preexistentes, contendo algumas anomalias definidas. Se uma nova anomalia for encontrada, o algoritmo proposto deve ser iniciado novamente para dar eficiência aos administradores da rede.

Se uma anomalia for detectada pelo algoritmo *Multilayer Perceptron*, deve-se relatar a intrusão ao administrador da rede, atualizando os registros em um arquivo de log. Caso contrário, retoma o fluxo normal de processamento.

C. Simuações e Resultados

Para realizar a validação do Sistema de Detecção de Intrusão proposto, o algoritmo K-Médias primeiro agrupa os dados em dez *clusters*. O algoritmo *Multilayer Perceptron*, por sua vez, realizará a classificação, fazendo uso das informações do *cluster*, resultante da etapa anterior, como entrada da Rede Neural, conforme Figura 5. Para a classificação a Rede Neural é treinada com 90% dados completo e 10% são utilizados para teste.

Para a realização do agrupamento dos dados define-se um total de 500 iterações e 10 *clusters* ou grupos para o algoritmo K-Médias. A rotulação das classes de anomalias para cada *cluster* é apresentado na Tabela I. A Figura 6 apresenta a discriminação dos dados em cada *cluster*, ou seja, apresenta a quantidade de registros de cada classe em cada *cluster*.

Para a classificação é utilizado um Conjunto de Testes, formado por 10% dos dados após o agrupamento dos mesmos.

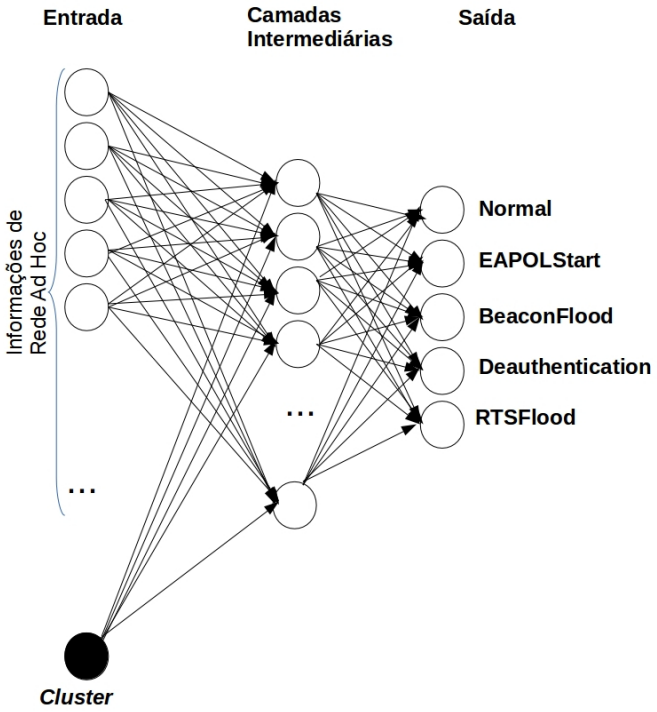


Figura 5. Proposta de Arquitetura da Rede Neural

Tabela I
ROTULAÇÃO DOS *Clusters*

Rotulação	Cluster
Normal	0,1,3,4,5,6,7,8,9
EAPOLStart	2
BeaconFlood	—
Deauthentication	—
RTSFlood	—

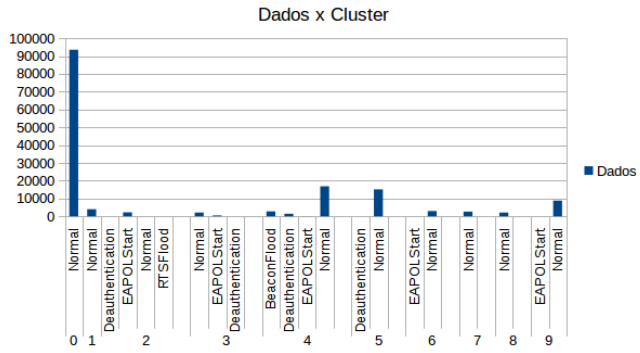


Figura 6. Dados por *Cluster*

O algoritmo *Multilayer Perceptron* irá treinar a Rede Neural através da base de treinamento, que é composta por 90% dos dados da Rede.

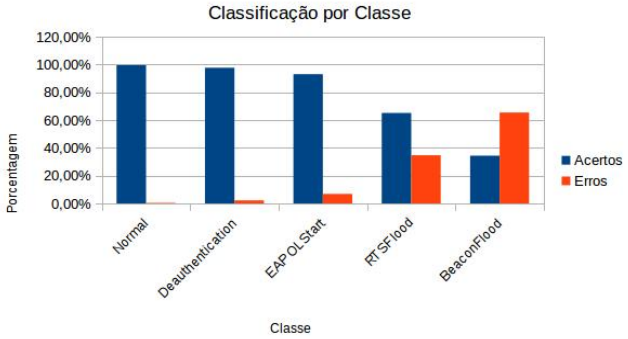


Figura 7. Classificação - *Multilayer Perceptron*

Tabela II
CLASSIFICAÇÃO DO SISTEMA PROPOSTO

Acertos	Erros
97%	3%

A Figura 7 apresenta o percentual de classificação em relação a cada classe através do algoritmo *Multilayer Perceptron*, que representa o percentual de dados classificados corretamente em cada classe. Enquanto que a Tabela II mostra a taxa de classificação total para o sistema proposto.

Os testes realizados no Sistema de Detecção de Intrusão proposto apontam para resultados relevantes. Para a validação do grupo de dados, são utilizados o banco de dados local [9] e o algoritmo K-Médias, que organiza os dados em 10 *clusters* em 97 iterações do algoritmo em 406,94 segundos. A organização dos dados pode ser verificada na Figura 6. Para a classificação dos dados já agrupados, é utilizado o algoritmo *Multilayer Perceptron*, que possui uma taxa de precisão de 97% dos dados, com erro médio em torno de 1,79% e erro quadrático médio em torno de 9,31%.

O sistema proposto é eficaz para o processo de classificação de dados de Redes Sem Fio Ad Hoc, através do uso dos algoritmos K-Médias e *Multilayer Perceptron*, pois permite a redução de falsos positivos, quando comparado ao uso isolado de estratégias de Inteligência Computacional adotadas.

Na Tabela III, é possível perceber que a proposta apresentada proporcionou uma taxa aceitável de sucesso, demonstrando que esta é uma abordagem viável para a construção do IDS.

VI. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresenta uma proposta de Sistema de Detecção de Intrusão em Redes Sem Fio Ad Hoc composto por duas etapas, baseado em agrupamento de dados através do algoritmo K-Médias e Redes Neurais Artificiais, para detecção e classificação de anomalias, causados por ataques às Redes de Computadores.

O algoritmo K-Médias permite agrupar os dados em grupos denominados *clusters*, que comporão dados com similaridade,

Tabela III
COMPARAÇÃO ENTRE DIVERSOS TRABALHOS

Proposta	**Classificação**
IDS Wavelet [8]	99%
IDS Híbrido [10]	98%
Nossa Proposta	97%
IDS Thatachi [3]	96%

através da distância euclidiana até o centróide mais próximo. Após este passo, o reconhecimento de padrões, que indica a anomalia, é simples e rápido. Para a detecção e classificação de novas anomalias, é necessário treinar novamente a Rede Neural.

O Sistema de Detecção de Intrusão proposto permite compartilhar as melhores características de cada método. A utilização em conjunto permite a redução de falso positivos, se comparado com a utilização isolada de ambas as técnicas.

A validação da proposta deste artigo fundamenta-se em dados obtidos de ambientes de Redes Sem Fio Ad Hoc doméstico e de organização. Para a classificação dos dados já agrupados utiliza-se o algoritmo *Multilayer Perceptron*, que possui taxa de acerto em 97% dos dados possuindo erro médio em torno de 1,79% e erro médio quadrático em torno de 9,31% em cada *cluster*. Os resultados obtidos aqui permitem concluir que a abordagem proposta é promissora, e um bom nıvel de detecção é conseguido nas avaliações realizadas.

Os trabalhos futuros podem ser realizados de diversas maneiras: aplicação da proposta em uma rede sem fio metropolitana, detecção online na rede. Também como trabalho futuro há a possibilidade de alteração do algoritmo para que execute de forma continua e que possa atualizar a Rede Neural para a detecção de novas anomalias.

REFERÊNCIAS

[1] Amudhavel, J., et al. *"A survey on intrusion detection system: State of the art review."* Indian Journal of Science and Technology 9.11 (2016).

[2] Brasil, P. *Brasil terminou janeiro com 257,248 milhões de acessos móveis.* http://www.brasil.gov.br/editoria/infraestrutura/2016/03/brasil-terminou-janeiro-com-257-248-milhoes-de-acessos-moveis. Accessed: 2018-06-01.

[3] Cervantes, Christian, Michele Nogueira, and Aldri Santos. *"Mitigação de Ataques no Roteamento em IoT Densa e Móvel Baseada em Agrupamento e Confiabilidade dos Dispositivos."* Simpósio Brasileiro de Redes de Computadores (SBRC). Vol. 36. 2018.

[4] Chandrashekar, Girish, and Ferat Sahin. *"A survey on feature selection methods."* Computers Electrical Engineering 40.1 (2014): 16-28.

[5] de Moraes, Rodrigo Fonseca, Nelcileno Virgílio de Souza Araújo, and Cristiano Maciel. *"Avaliação de um Conjunto de Dados Quanto à sua Qualidade na Especificação de Perfis de Ataque e Não-Ataque Numa Rede IEEE 802.11 w."* Anais da Escola Regional de Informática da Sociedade Brasileira de Computação (SBC)–Regional de Mato Grosso 6 (2015): 145-150.

[6] Dorri, Ali, Seyed Reza Kamel, and Esmaeil Kheirkhah. *"Security challenges in mobile ad hoc networks: a survey."* arXiv preprint arXiv:1503.03233 (2015).

[7] Ferreira, Ed' Wilson Tavares. *Proposta de um sistema de detecção e classificação de intrusão em redes de computadores baseado em transformadas wavelets e redes neurais artificiais*. Tese (doutorado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Engenharia Elétrica - 2009.

[8] Ferreira, Ed Wilson Tavares, et al. *"Intrusion detection system with wavelet and neural artifical network approach for networks computers."* IEEE Latin America Transactions 9.5 (2011): 832-837.

[9] Ferreira, Ed Wilson Tavares, et al. *"A Methodology for building a Dataset to Assess Intrusion Detection Systems in Wireless Networks."* WSEAS Transactions on Communications, 14 (2015) 113-120.

[10] Govindarajan, M., and R. M. Chandrasekaran. *"Intrusion detection using neural based hybrid classification methods."* Computer networks 55.8 (2011): 1662-1671.

[11] Haq, Nutan Farah, et al. *"Application of machine learning approaches in intrusion detection system: a survey."* IJARAI-International Journal of Advanced Research in Artificial Intelligence 4.3 (2015): 9-18.

[12] Haykin, Simon S., et al. *Neural networks and learning machines.* Vol. 3. Upper Saddle River: Pearson, 2009.

[13] Loo, Jonathan, Jaime Lloret Mauri, and Jesus Hamilton Ortiz, eds. *Mobile ad hoc networks: current status and future trends.* CRC Press, 2016.

[14] MacQueen, James. *"Some methods for classification and analysis of multivariate observations."* Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1. No. 14. 1967.

[15] Marins, Aldecir Xavier de. *"Protocolos de roteamento para redes móveis comparativo: OLSR X AODV."* (2017).

Daniel Rosa Canêdo possui graduação em Engenharia de Computação pela Pontifícia Universidade Católica de Goiás (2003) e com mestrado em Engenharia Elétrica pela Universidade de Brasília (2006). Atualmente é professor exclusivo do Instituto Federal de Goiás - Campus Luziânia. Atualmente é aluno de doutorado do Programa de Pós-Graduação em Engenharia de Sistemas Eletrônicos e Automação do Departamento de Engenharia Elétrica da Universidade de Brasília (UnB).

Alexandre Ricardo Soares Romariz possui graduação em Engenharia Elétrica pela Universidade de Brasília (1992), Mestre em Engenharia Elétrica pela Universidade Estadual de Campinas (1995) e Doutor em Engenharia Elétrica pela Universidade do Colorado em Boulder (2003). Atualmente é professor associado na Universidade de Brasília. Tem experiência na área de Inteligência Computacional, Circuitos Integrados, Optoeletrônica e Processamento Digital de Sinais.